

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: February 11, 2003, 11:46:34 ; Search time 28.2264 Seconds  
(without alignments)  
1605.064 Million cell updates/sec

Title: US-10-021-121-4

Perfect score: 1850

Sequence: 1 MGPPIHSGPGRVRCALLLG.....PYIVQDGPQSPPIRYKV 340

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 50 summaries

Database :

A\_Geneseq\_101002:\*  
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1850	100.0	340	AAW33699	AL-2-short (AL-2s)
2	1850	100.0	340	AAW31544	Human cytokine Lcr
3	1850	100.0	340	AAW10637	NLRK2 ligand for
4	1844	99.7	340	AAW17081	EPH family ligand
5	1841	99.5	455	AAW33698	AL-2-long (AL-2l)
6	1836	99.2	340	AAW46615	Human transmembran
7	637.5	34.5	334	AAW00287	Mouse Eph receptor
8	637.5	34.5	336	AAW32742	Murine hepatoma tr
9	632	34.6	346	AAW55059	Elk tyrosine kinase
10	632	34.2	346	AAW91930	Human cytokine elk

11	632	34.2	346	AAW6055	Human elk-L protei
12	632	34.2	346	AAW19249	Human elk ligand p
13	632	34.2	346	AAW44233	Human elk-L. Homo
14	631	34.1	346	AAW82606	Eph transmembrane
15	630.5	34.0	331	AAW00288	Chicken Eph recept
16	629.5	34.0	333	AAW06337	Full length ligand
17	629.5	34.0	333	AAW4655	Ligand for recepto
18	629.5	34.0	333	AAW32743	Human hepatoma tra
19	629.5	34.0	333	AAW8287	Human LERK-5 cyto
20	629.5	34.0	333	AAW11308	Receptor-type tyro
21	619.5	33.5	308	AAW06334	Ligand #2 for rece
22	619.5	33.5	308	AAW4656	Ligand for recepto
23	492	26.6	89	AAW11438	Human ephrin B3 C-
24	453.5	24.5	658	AAW6782	Ephrin-B1-Ephrin-B
25	453.5	24.5	683	AAW6781	Murine ephrin B2.1
26	447	24.2	229	AAW24019	Ligand #1 for rece
27	443	23.9	229	AAW24020	Receptor-type tyro
28	431.5	23.3	195	AAW06333	Ligand #1 for rece
29	431.5	23.3	195	AAW11307	Receptor-type tyro
30	284.5	15.4	92	AAW37671	Peptide #11708 enc
31	284.5	15.4	92	ABG46524	Human peptide enco
32	205.5	11.1	136	AAW37534	Peptide #11571 enc
33	205.5	11.1	136	ABG46394	Human peptide enco
34	204	11.0	82	AAW11437	Human ephrin B2 C-
35	202.5	10.9	106	AAW54187	Human pancreatic C
36	201	10.9	82	AAW11436	Human ephrin B1 C-
37	182	9.8	652	ABW5548	Drosophila melanog
38	182	9.8	652	ABW5548	Eph transmembrane
39	181	9.8	234	AAW82605	Human hek-L protei
40	179	9.7	238	AAW11481	Leuk-6 protein. M
41	176	9.5	184	AAW02587	Leuk-6 protein. M
42	176	9.5	184	AAW1283	Murine LERK-6 poly
43	176	9.5	184	AAW06820	Mouse EPH receptor
44	176	9.5	209	AAW94766	Human acid sequenc
45	176	9.5	209	AAW1006	Human LERK-6 poly
46	175.5	9.5	213	AAW06822	Human pancreatic C
47	172.5	9.3	218	AAW54058	Chicken EPH recept
48	172	9.3	200	AAW94767	Amino acid sequenc
49	172	9.3	200	AAW11007	Human hek-L protei
50	169.5	9.2	201	AAW11482	

#### ALIGNMENTS

RESULT 1  
AAW33699 standard; Protein: 340 AA.

AC AAW33699;

DT 30-APR-1998 (first entry)

XX AL-2-short (AL-2s) protein.

XX AL-2s; AL-2; AL-2-short; human; treatment; neurological disorder; tumour;

KM rheumatoid arthritis; wound healing; paralysis; angiogenesis; leukaemia;

XX psoriasis; Alzheimer's disease; epilepsy.

OS Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

Location/Qualifiers  
1..26  
/note="signal peptide"  
27..340  
/note="mature protein"  
27..219  
/note="extracellular domain"  
220..245  
/note="hydrophobic transmembrane domain"

W09740153-AL.

PD 30-OCT-1997.  
 XX 17-APR-1997; 97WO-US06345.  
 PF 19-APR-1996; 96US-0635130.  
 PR (GERTH ) GENENTECH INC.  
 PA Caras IW;  
 XX WPI; 1997-535837/49.  
 XX N-PSDB; AAV06355.  
 DR Human AL-2 neurotrophic factor and related DNA - used to develop  
 PT products for, e.g. treating neurologic disorders, angiogenesis  
 PT disorders, tumours or rheumatoid arthritis or for wound healing  
 XX  
 PS Claim 20; Fig 2A-B; 86pp; English.

CC This is a AL-2-short (AL-2s) protein. The AL-2 is a novel Eph-related  
 CC tyrosine kinase receptor ligand. AL-2 can be administered to patients in  
 CC whom the nervous system has been damaged by trauma, surgery, stroke,  
 CC ischaemia, infection, metabolic disease, nutritional deficiency,  
 CC malignancy, or toxic agents, to promote the survival or growth of  
 CC neurons. They can be used to treat motoneuron disorders such as  
 CC amyotrophic lateral sclerosis (Lou Gehrig's disease), Bell's palsy, and  
 CC various conditions involving spinal muscular atrophy, or paralysis. AL-2  
 CC can be used to treat human neurodegenerative disorders, such as  
 CC Alzheimer's disease, Parkinson's disease, epilepsy, demyelinating  
 CC diseases such as multiple sclerosis, Huntington's chorea, Down's syndrome,  
 CC nerve deafness, Menier's disease, and other disorders of the cerebellum.  
 CC AL-2 can be used as cognitive enhancer, to enhance learning particularly  
 CC in dementias or trauma, since they can promote axonal outgrowth and  
 CC synaptic plasticity, particularly of hippocampal neurons that express  
 CC AL-2 binding Eph-family receptors and cortical neurons that express  
 CC AL-2. AL-2 can also be used for wound healing, i.e. accelerating  
 CC neovascularisation of, e.g. burns and ulcers. The encoding nucleic acids  
 CC are useful in preparing antibodies that specifically bind to the AL-2  
 CC protein. The antibodies and the AL-2 antagonists are useful in diagnosing  
 CC and treating various neuronal disorders. AL-2 antagonists can be used  
 CC for modulating angiogenesis. They can also be used for the treatment of  
 CC tumours, acute myeloid leukaemia (AML), chronic myeloid leukaemia (CML),  
 CC myelodysplastic syndrome (MDS), diabetic retinopathy, neovascular  
 CC glaucoma, psoriasis and rheumatoid arthritis.

Sequence 340 AA:

Query Match 100.0%; Score 1850; DB 18; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-143;  
 Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPFHGPGVVRGALLLLGLVGLVGLSLPEYVWNSANKRFQAGGYLYVQIGRDL 60  
 DB 1 MGPFHGPGVVRGALLLLGLVGLVGLSLPEYVWNSANKRFQAGGYLYVQIGRDL 60  
 QY 61 LCPRRAPPGPHSSPNYEFKLYLVGAGGRRCCEAPPANLLTCRPPDIDLEFTTKFOEY 120  
 DB 61 LCPRRAPPGPHSSPNYEFKLYLVGAGGRRCCEAPPANLLTCRPPDIDLEFTTKFOEY 120  
 QY 121 SPVLWHEPFSHHYITLTSOSTREGLSLOGGVCLTGMVLYLRVQSPGGAVPKRP 180  
 DB 121 SPVLWHEPFSHHYITLTSOSTREGLSLOGGVCLTGMVLYLRVQSPGGAVPKRP 180  
 QY 181 VSEMPERDRGAHSLPEKENVLPDPTSNATSRGAGBLPPSPMPAVAGAGLALLLL 240  
 DB 181 VSEMPERDRGAHSLPEKENVLPDPTSNATSRGAGBLPPSPMPAVAGAGLALLLL 240  
 QY 241 GVAGAGAMWRRRRAKPPSSSRPPGSGRGSLGLGGCGMGPREAPGLGIALRGG 300  
 DB 241 GVAGAGAMWRRRRAKPPSSSRPPGSGRGSLGLGGCGMGPREAPGLGIALRGG 300  
 QY 301 GAADPPFCPHEKVGSDYGHVYIVDGGPQSPFNIIYRV 340

DB 301 GAADPPFCPHEKVGSDYGHVYIVDGGPQSPFNIIYRV 340

RESULT 2  
 ID AAW31544  
 AC AAW31544 standard; Protein; 340 AA.  
 XX AAW31544;  
 DT 14-APR-1998 (first entry)  
 XX Human cytokine Lerk-8.  
 DE Lerk-8; cytokine; human; hek; elk; receptor tyrosine kinase;  
 KW ligand; neurodegenerative disease; wound healing;  
 KM neovascularisation; diagnosis; therapy.  
 XX Homo sapiens.

FT Key Location/Qualifiers  
 FT Peptide 1..27  
 FT Protein /label= Sig\_peptide  
 FT Protein 28..340  
 FT Domain /label= Mat\_protein  
 FT Domain 28..224  
 FT Domain /note= "extracellular domain"  
 FT Domain 225..251  
 FT Domain /note= "transmembrane domain"  
 FT Domain 252..340  
 FT Modified-site /note= "cytoplasmic domain"  
 FT Modified-site 210..212  
 FT /note= "N-glycosylated"  
 FT MISC-difference 325  
 FT /note= "residue 325 is Leu in Lerk-8 variant"

PN WO9736919-A2.  
 XX 09-OCT-1997.  
 PD 19-MAR-1997; 97WO-US04533.  
 XX 21-MAR-1996; 96US-0621146.  
 PR (IMMV ) IMMUNEX CORP.  
 PA Cerrectl DP;  
 PI WPI; 1997-503043/46.  
 DR N-PSDB; AAT89519.

PT New isolated cytokine, Lerk-8 - binds to the hek and elk receptor  
 PT tyrosine kinases, used to develop products for diagnosis and therapy  
 XX  
 PS Claim 3; Page 32-33; 37pp; English.

CC This protein sequence comprises a novel human cytokine designated  
 CC Lerk-8. The amino acid sequence was deduced from a human foetal  
 CC brain cDNA clone (see AAT89519). Lerk-8 binds to the cell surface  
 CC receptors hek and elk, which are members of the eph/elk family of  
 CC receptor tyrosine kinases. Lerk-8 polypeptides, especially soluble  
 CC polypeptides comprising amino acid residues -27 to 142-197 of the  
 CC full-length protein, can be expressed in transformed host cells.  
 CC These polypeptides can be used to purify hek or elk proteins, and  
 CC to purify or identify cells that express hek or elk on the surface.  
 CC Such cells can be used in various in vitro studies or in vivo  
 CC procedures, e.g. neural cells expressing elk can be administered to  
 CC a mammal afflicted with a neurodegenerative disorder. The Lerk-8  
 CC polypeptides can also be used to deliver diagnostic or therapeutic  
 CC agents to these cells (e.g. leukaemia cells). The Lerk-8 DNA and  
 CC polypeptides can also be used to: treat disorders mediated by  
 CC defective or insufficient amounts of Lerk-8; to treat disorders  
 CC such as injury to neural tissue or neurologic disease; to promote  
 CC angiogenesis; and for wound healing or stimulating

CC neovascularisation of grafted tissues.  
XX  
SQ Sequence 340 AA;

Query Match 100.0%; Score 1850; DB 18; Length 340;  
Best Local Similarity 100.0%; Pred. No. 2.9e-143;  
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPPHSGGVRVAGALLLLGVLGIVSGLSLEPYWNSANKRFOAEGGYVLYPQIGDRIDL 60  
DB 1 MGPPHSGGVRVAGALLLLGVLGIVSGLSLEPYWNSANKRFOAEGGYVLYPQIGDRIDL 60  
QY 61 LCPRARPPGPHSSPNVEFYKLYLVGAQGRCEAPPAENLLITCDRBDLDLRTIKFOEY 120  
DB 61 LCPRARPPGPHSSPNVEFYKLYLVGAQGRCEAPPAENLLITCDRBDLDLRTIKFOEY 120  
QY 121 SPNLMGHEFRSHHDYIIATSDGTREGLESLOGVCCTTRGMKVLRLVQSPRGAVPRKP 180  
DB 121 SPNLMGHEFRSHHDYIIATSDGTREGLESLOGVCCTTRGMKVLRLVQSPRGAVPRKP 180  
QY 181 VSEWPMERDRGAASLEPGKENTLPGDPTSNATSRGAEGPLPPSPMPAVAGAAGLALLLL 240  
DB 181 VSEWPMERDRGAASLEPGKENTLPGDPTSNATSRGAEGPLPPSPMPAVAGAAGLALLLL 240  
QY 241 GVAAGAGAMCWRRRARAPSRBSRHPGSGFGRGSLGLGGGGGMPREAPBEGELIALRG 300  
DB 241 GVAAGAGAMCWRRRARAPSRBSRHPGSGFGRGSLGLGGGGGMPREAPBEGELIALRG 300  
QY 301 GAADPPPCPHYEKVSGDYGHVYIVODGPPQSPNNITYKV 340  
DB 301 GAADPPPCPHYEKVSGDYGHVYIVODGPPQSPNNITYKV 340

RESULT 3  
AAW10637  
ID AAW10637 standard; Protein; 340 AA.

XX AAW10637;  
XX 23-JUN-1997 (first entry)  
XX  
XX NLERK2 ligand for eph-related kinase.  
XX  
XX LERK: ligand for eph-related kinase; ERK; NLERK2;  
XX receptor protein tyrosine kinase; cell proliferation;  
XX cell differentiation; cell survival; nerve cell.  
XX  
XX Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..29  
FT /label= Sig\_peptide  
FT Protein 30..340  
FT /label= Mat\_protein  
FT Domain 227..251  
FT /label= Transmembrane\_domain  
FT Modified-site 210  
FT /label= N-glycosylation\_site  
XX  
XX W09704091-A1.  
XX  
XX 06-FEB-1997.  
XX  
XX 19-JUL-1996; 96WO-AU00460.  
XX  
XX 05-FEB-1996; 96AU-0007890.  
XX 20-JUL-1995; 95AU-0004263.  
XX 27-NOV-1995; 95AU-0006847.  
XX 22-DEC-1995; 95AU-0007299.  
XX  
XX (AMRA-) AMRAD OPERATIONS PTY LTD.  
XX  
XX Nicola NA;  
FI

XX WPI, 1997-132632/12.  
DR N-PSDB; AAT60966.  
XX

PT Nucleic acid mol. encoding ligand for eph-related kinase - useful  
FT for treatment of, pref. neuronal, cells to increase survival,  
PT proliferation and differentiation  
XX

PS Claim 16; Page 37-39; 71pp; English.

XX A novel human ligand for eph-related kinase (LERK) is designated  
CC NLERK2 (AAW10637). It is encoded by a cDNA clone (AAT60966) obtd.  
CC from a human foetal brain cDNA library. The novel receptor ligand  
CC can be expressed in transformed host cells and used in methods  
CC for regulating the development, maintenance or regeneration of  
CC different cells (e.g. neurons) and tissues in vivo and in vitro.  
CC Soluble NLERK2 peptides can be used to treat injury, disease or  
CC abnormality in the nervous system, and membrane-bound NLERK2 to  
CC modulate proliferation, different or survival e.g. in grafting  
CC procedures or transplantation. NLERK2 can also be used to raise  
CC antibodies for use in immunotherapy, and to detect anti-NLERK2  
CC antibodies that may occur in some autoimmune diseases.

XX Sequence 340 AA;

Query Match 100.0%; Score 1850; DB 18; Length 340;  
Best Local Similarity 100.0%; Pred. No. 2.9e-143;  
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPPHSGGVRVAGALLLLGVLGIVSGLSLEPYWNSANKRFOAEGGYVLYPQIGDRIDL 60  
DB 1 MGPPHSGGVRVAGALLLLGVLGIVSGLSLEPYWNSANKRFOAEGGYVLYPQIGDRIDL 60  
QY 61 LCPRARPPGPHSSPNVEFYKLYLVGAQGRCEAPPAENLLITCDRBDLDLRTIKFOEY 120  
DB 61 LCPRARPPGPHSSPNVEFYKLYLVGAQGRCEAPPAENLLITCDRBDLDLRTIKFOEY 120  
QY 121 SPNLMGHEFRSHHDYIIATSDGTREGLESLOGVCCTTRGMKVLRLVQSPRGAVPRKP 180  
DB 121 SPNLMGHEFRSHHDYIIATSDGTREGLESLOGVCCTTRGMKVLRLVQSPRGAVPRKP 180  
QY 181 VSEWPMERDRGAASLEPGKENTLPGDPTSNATSRGAEGPLPPSPMPAVAGAAGLALLLL 240  
DB 181 VSEWPMERDRGAASLEPGKENTLPGDPTSNATSRGAEGPLPPSPMPAVAGAAGLALLLL 240  
QY 241 GVAAGAGAMCWRRRARAPSRBSRHPGSGFGRGSLGLGGGGGMPREAPBEGELIALRG 300  
DB 241 GVAAGAGAMCWRRRARAPSRBSRHPGSGFGRGSLGLGGGGGMPREAPBEGELIALRG 300  
QY 301 GAADPPPCPHYEKVSGDYGHVYIVODGPPQSPNNITYKV 340  
DB 301 GAADPPPCPHYEKVSGDYGHVYIVODGPPQSPNNITYKV 340

RESULT 4  
AAW17081  
ID AAW17081 standard; Protein; 340 AA.

XX AAW17081;  
XX 09-AUG-1997 (first entry)  
XX  
XX Eph family ligand Efl-6.  
XX  
XX Efl-6; Eph; Elk; receptor tyrosine kinase; signal transduction;  
XX ligand; neurological disease.  
XX  
XX Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..24  
FT /label= Sig\_peptide  
FT Protein 25..340

```

FT FT /label= Mat_protein
FT Domain 225..249
FT /label= Transmembrane_domain
FT Misc-difference 166
FT /label= Gln, Arg
XX
XX WO9715667-A1.
XX
XX 01-MAY-1997.
XX
XX 25-OCT-1996; 96WO-US17201.
XX
XX 25-OCT-1995; 95US-0007015.
XX
XX (REGG-) REGENERON PHARM INC.
XX
XX Davis S, Gale NW, Yancopoulos GD;
XX
XX WPI: 1997-259021/23.
XX
XX N-PSDB; AA69808.
XX
XX New nucleic acid encoding Efl-6 ligand protein - used for promoting
XX growth and proliferation of neuronal cells and in drug screening
XX
XX Claim 2; Fig 1; 36pp; English.
XX
XX A novel ligand (AAW17081), designated Efl-6 (or Eph transmembrane
XX tyrosine kinase family ligand 6), binds to the Elk, Nuk/Cek5,
XX HeK2/Seq4, Htk and Sek1 receptors on cells. Its amino acid
XX sequence was deduced from a human frontal cortex cDNA clone
XX (AA69808). Recombinant Efl-6, truncated soluble polypeptides
XX comprising the extracellular domain of Efl-6, and Efl-6
XX ligandbodies comprising soluble Efl-6 and the FC portion of IgG can
XX be expressed in host cells. These can be used to support neuronal
XX and other Eph receptor-bearing cell populations for treatment of
XX neurological disorders, in drug screening and to raise diagnostic
XX antibodies.
XX
XX Sequence 340 AA;
SQ
Query Match 99.7%; Score 1844; DB 18; Length 340;
Best Local Similarity 99.7%; Pred. No. 9e-143;
Matches 339; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 MGPFGSGPGVRYGALLLVGLVGLVSLLEPYWNSANKRFOAEGVLYPQIDRLD, 60
XX 1 MGPFGSGPGVRYGALLLVGLVGLVSLLEPYWNSANKRFOAEGVLYPQIDRLD, 60
XX 1 MGPFGSGPGVRYGALLLVGLVGLVSLLEPYWNSANKRFOAEGVLYPQIDRLD, 60
XX
XX 61 LCPRARPPGPHSSPNVEFYKLYVGAQGRCEAPAPMLLTCDRPDLNFTTKFOEY 120
XX 61 LCPRARPPGPHSSPNVEFYKLYVGAQGRCEAPAPMLLTCDRPDLNFTTKFOEY 120
XX 61 LCPRARPPGPHSSPNVEFYKLYVGAQGRCEAPAPMLLTCDRPDLNFTTKFOEY 120
XX
XX 121 SPFLMGHEPFSHDDYITISDTRGSLGQGVCLTRGMKYLRVGQSPRGAVPRXP 180
XX 121 SPFLMGHEPFSHDDYITISDTRGSLGQGVCLTRGMKYLRVGQSPRGAVPRXP 180
XX 121 SPFLMGHEPFSHDDYITISDTRGSLGQGVCLTRGMKYLRVGQSPRGAVPRXP 180
XX
XX 181 VSMEMERDRGAASLEPEKENTPDPPTSNATSRGAEGLPPSPMPAVAGAAIGLALLL 240
XX 181 VSMEMERDRGAASLEPEKENTPDPPTSNATSRGAEGLPPSPMPAVAGAAIGLALLL 240
XX 181 VSMEMERDRGAASLEPEKENTPDPPTSNATSRGAEGLPPSPMPAVAGAAIGLALLL 240
XX
XX 241 GVGAGAGAMCMRRRAKPSRSRHPGSGFRGSGSLGAGGGGAGPEAPGELIALRGG 300
XX 241 GVGAGAGAMCMRRRAKPSRSRHPGSGFRGSGSLGAGGGGAGPEAPGELIALRGG 300
XX 241 GVGAGAGAMCMRRRAKPSRSRHPGSGFRGSGSLGAGGGGAGPEAPGELIALRGG 300
XX
XX 301 GAADPPFCFHYEKVSGDYGHVYIVDDGPPQSPPNITYKY 340
XX 301 GAADPPFCFHYEKVSGDYGHVYIVDDGPPQSPPNITYKY 340
XX 301 GAADPPFCFHYEKVSGDYGHVYIVDDGPPQSPPNITYKY 340
XX
RESULT 5
AAW33698
ID AAW33698 standard; Protein; 455 AA.
XX

```

```

AC AAW33698;
XX
XX 30-APR-1998 (first entry)
XX
XX AL-2-long (AL-21) protein.
XX
XX AL-21; AL-2; AL-2-long; human; treatment; neurological disorder; tumour;
XX rheumatoid arthritis; wound healing; paralysis; angiogenesis; leukaemia;
XX psoriasis; Alzheimer's disease; epilepsy.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 1..26
XX Protein /note= "signal peptide"
XX /note= "mature protein"
XX Domain 27..455
XX /note= "extracellular domain"
XX /note= "220..245
XX Domain /note= "hydrophobic transmembrane domain"
XX
XX WO9740153-A1.
XX
XX 30-OCT-1997.
XX
XX 17-APR-1997; 97WO-US06345.
XX
XX 19-APR-1996; 96US-0635130.
XX
XX (GETH ) GENENTECH INC.
XX
XX Caras IW;
XX
XX WPI: 1997-535837/49.
XX
XX N-PSDB; AAV06354.
XX
XX Human AL-2 neurotrophic factor and related DNA - used to develop
XX products for, e.g. treating neurologic disorders, angiogenesis
XX disorders, tumours or rheumatoid arthritis or for wound healing
XX
XX Claim 20; Fig 1A-C; 86pp; English.
XX
XX This is a AL-2-long (AL-21) protein. The AL-2 is a novel Eph-related
XX tyrosine kinase receptor ligand. AL-2 can be administered to patients in
XX whom the nervous system has been damaged by trauma, surgery, stroke,
XX ischaemia, infection, metabolic disease, nutritional deficiency,
XX malignancy, or toxic agents, to promote the survival or growth of
XX neurons. They can be used to treat motorneuron disorders such as
XX amyotrophic lateral sclerosis (Lou Gehrig's disease), Bell's palsy, and
XX various conditions involving spinal muscular atrophy, or paralysis. AL-2
XX can be used to treat human neurodegenerative disorders, such as
XX Alzheimer's disease, Parkinson's disease, epilepsy, demyelinating
XX diseases such as multiple sclerosis, Huntingtons chorea, Down's syndrome,
XX nerve deafness, Menier's disease, and other disorders of the cerebellum.
XX AL-2 can be used as cognitive enhancer, to enhance learning particularly
XX in dementia or trauma, since they can promote axonal outgrowth and
XX synaptic plasticity, particularly of hippocampal neurons that express
XX AL-2 binding Eph-family receptors and cortical neurons that express
XX AL-2. AL-2 can also be used for wound healing, i.e. accelerating
XX neovascularisation of, e.g. burns and ulcers. The encoding nucleic acids
XX are useful in preparing antibodies that specifically bind to the AL-2
XX protein. The antibodies and the AL-2 antagonists are useful in diagnosing
XX and treating various neuronal disorders. AL-2 antagonists can be used
XX for modulating angiogenesis. They can also be used for the treatment of
XX tumours, acute myeloid leukaemia (AML), chronic myeloid leukaemia (CML),
XX myelodysplastic syndrome (MDS), diabetic retinopathy, neovascular
XX glaucoma, psoriasis and rheumatoid arthritis.
XX
XX Sequence 455 AA;
SQ
Query Match 99.5%; Score 1841; DB 18; Length 455;
Best Local Similarity 100.0%; Pred. No. 2.2e-142;

```



Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MGPHSGGGRVAGALLLGLVLSGLSLPEVYWNANRFOAEGGYLYPQIGDRDL 60
DB 1 MGPHSGGGRVAGALLLGLVLSGLSLPEVYWNANRFOAEGGYLYPQIGDRDL 60
QY 61 LCPRARPPGPHSSPNVEFYKLYLVGAQGRCEAPAPNLLTCDRDPDLRFITKFOEY 120
DB 61 LCPRARPPGPHSSPNVEFYKLYLVGAQGRCEAPAPNLLTCDRDPDLRFITKFOEY 120
QY 121 SPNMGHEFRSHHDYIYIATSDGTREGLESIOGGVCLTRGMKVLLRVGQSPRGAVPRKP 180
DB 121 SPNMGHEFRSHHDYIYIATSDGTREGLESIOGGVCLTRGMKVLLRVGQSPRGAVPRKP 180
QY 181 VSEMPMERDRGAASLEPGKENTLPGDPTSNATSRGABGLPPSPMPAVAGAAGLALLL 240
DB 181 VSEMPMERDRGAASLEPGKENTLPGDPTSNATSRGABGLPPSPMPAVAGAAGLALLL 240
QY 241 GVAAGAGAMCWRRRRAKPSRRHPGSGFGRGSLGLGGGGMGPREAPBELGIALRG 300
DB 241 GVAAGAGAMCWRRRRAKPSRRHPGSGFGRGSLGLGGGGMGPREAPBELGIALRG 300
QY 301 GAADPPFCPHYEKVSGDYGHPIYIVDGPQSPNITY 338
DB 301 GAADPPFCPHYEKVSGDYGHPIYIVDGPQSPNITY 338

```

## RESULT 6

AAW46615  
ID AAW46615 standard; Protein; 340 AA.

AC AAW46615;  
XX  
DT 06-JUL-1998 (first entry)  
XX

DE Human transmembrane ligand Elk-L3.

XX Elk-L3; Elk-related receptor tyrosine kinase; transmembrane ligand;  
KW human; signal transduction; axonogenesis; nerve cell; neurons;  
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;  
KW demyelination; multiple sclerosis; amyotrophic lateral sclerosis;  
KW nervous system infection; Wernicke's disease; trauma; ischaemia;  
KW stroke; nutritional polyneuropathy; progressive supranuclear palsy;  
KW Shy Drager's syndrome; multistem degeneration;  
KW olivoponto cerebellar atrophy; peripheral nerve damage.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 166 /label= Gln, Arg

FT Domain 225..249 /note="transmembrane domain"

XX MO9801548-A1.

XX 15-JAN-1998.

XX 04-JUL-1997; 97MO-CA00473.

XX 05-JUL-1996; 96US-0021272.

XX (MOUN) MOUNT SINAI HOSPITAL CORP.

XX PA Holland S, Mdamaju G, Dawson T;

XX WPI; 1998-101047/09.

XX DR N-PSDB; AAV16097.

XX Modulating transmembrane ligand for an Elk-related receptor tyrosine

PT kinase - by formation of a complex between an oligomerised

PT Elk-related receptor tyrosine kinase and a transmembrane ligand

XX

PS Disclosure; Fig 5A; 40pp; English.

XX This polypeptide comprises human Elk-L3, a transmembrane ligand of  
CC Elk-related receptor tyrosine kinase (ERK). A novel method of  
CC modulating the biological activity of, or for affecting a pathway  
CC regulated by, a transmembrane ligand for an ERK in a cell  
CC expressing the transmembrane ligand comprises forming a complex  
CC between a purified and isolated oligomerised ERK, or an isoform  
CC or an extracellular domain of the ERK, and the transmembrane  
CC ligand expressed on the cell. The complex can also be used for  
CC evaluating a substance for its ability to modulate the biological  
CC activity of a transmembrane ligand for an ERK, and to identify  
CC substances that affect or modulate a pathway regulated by a ERK.  
CC A purified and isolated oligomerised ERK can be used in the  
CC preparation of a medicament for modulating neuronal development or  
CC regeneration in a subject, or in a medicament for modulating  
CC axonogenesis in a subject (all claimed). The substances identified  
CC by the methods can be used to modulate axonogenesis, nerve cell  
CC interactions and regeneration, to treat diseases and conditions  
CC involving trauma and injury to the nervous system, such as  
CC Alzheimer's disease, Parkinson's disease, Huntington's disease,  
CC demyelinating diseases, such as multiple sclerosis, amyotrophic  
CC lateral sclerosis, bacterial and viral infections of the nervous  
CC system, deficiency diseases, such as Wernicke's disease and  
CC nutritional polyneuropathy, progressive supranuclear palsy,  
CC Shy Drager's syndrome, multistem degeneration and olivoponto  
CC cerebellar atrophy, peripheral nerve damage, trauma, and ischaemia  
CC resulting from stroke.

XX Sequence 340 AA;

Query Match 99.2%; Score 1836; DB 19; Length 340;

Best Local Similarity 99.4%; Pred. No. 4e-142;

Matches 338; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 MGPHSGGGRVAGALLLGLVLSGLSLPEVYWNANRFOAEGGYLYPQIGDRDL 60
DB 1 MGPHSGGGRVAGALLLGLVLSGLSLPEVYWNANRFOAEGGYLYPQIGDRDL 60
QY 61 LCPRARPPGPHSSPNVEFYKLYLVGAQGRCEAPAPNLLTCDRDPDLRFITKFOEY 120
DB 61 LCPRARPPGPHSSPNVEFYKLYLVGAQGRCEAPAPNLLTCDRDPDLRFITKFOEY 120
QY 121 SPNMGHEFRSHHDYIYIATSDGTREGLESIOGGVCLTRGMKVLLRVGQSPRGAVPRKP 180
DB 121 SPNMGHEFRSHHDYIYIATSDGTREGLESIOGGVCLTRGMKVLLRVGQSPRGAVPRKP 180
QY 181 VSEMPMERDRGAASLEPGKENTLPGDPTSNATSRGABGLPPSPMPAVAGAAGLALLL 240
DB 181 VSEMPMERDRGAASLEPGKENTLPGDPTSNATSRGABGLPPSPMPAVAGAAGLALLL 240
QY 241 GVAAGAGAMCWRRRRAKPSRRHPGSGFGRGSLGLGGGGMGPREAPBELGIALRG 300
DB 241 GVAAGAGAMCWRRRRAKPSRRHPGSGFGRGSLGLGGGGMGPREAPBELGIALRG 300
QY 301 GAADPPFCPHYEKVSGDYGHPIYIVDGPQSPNITYKV 340
DB 301 GAADPPFCPHYEKVSGDYGHPIYIVDGPQSPNITYKV 340

```

## RESULT 7

AAW00287  
ID AAW00287 standard; Protein; 334 AA.

XX AAW00287;

XX 19-JAN-1997 (first entry)

XX Mouse Eph receptor ligand ELF-2.

XX Eph receptor; ligand; ELF-2; tyrosine kinase; signal transduction;

KW organogenesis; oncogenesis; tumour; neurological disorder;

KW diagnosis; gene therapy.

[illegible]

```

RESULT 8
AAR92742 ID AAR92742 standard; Protein: 336 AA.
XX AAR92742;
XX AAR92742;
XX 21-MAY-1996 (first entry)
XX Murine hepatoma transmembrane kinase receptor ligand.
XX Hepatoma transmembrane kinase; Htk; receptor; ligand;
XX Tyrosine kinase; neurodegenerative disease.
XX Mus musculus.
XX WO9602645-A2.
XX 01-FEB-1996.
XX 14-JUL-1995; 95WO-US08812.
XX 20-JUL-1994; 94US-0277722.
XX (GETH ) GENENTECH INC.
XX Bennett BD, Matthews W;
XX WPI: 1996-105907/11.
XX N-PSDB; AAT16470.
XX Ligand for the hepatoma trans-membrane kinase receptor - useful for
XX stimulating and inhibiting cells carrying the receptor, e.g. for
XX treating neuro-degenerative disease
XX Claim 5; Fig 1(A-D); 88pp; English.
XX Mouse (AAT16470) and human (AAT16471) Htk ligand which bind to, and
XX activate, the Htk receptor, have been identified in a variety of
XX tissues using a soluble Htk-Fc fusion protein.
XX The predicted mol.wt. of the murine Htk ligand protein following a
XX signal peptide cleavage is 34 kD with an estimated pI of 8.9.
XX The murine and human ligands show 96% homology at the amino acid
XX level.
XX The DNA is used to produce recombinant ligands; for tissue-
XX specific typing (partic. as a marker for breast cancer) and as a
XX marker for human chromosome 13. The ligands (partic. in soluble
XX form) are used to activate the tyrosine kinase domain of the
XX Htk receptor, i.e. to stimulate or inhibit growth, differentiation,
XX and/or activation of cells contg. the receptor, e.g. treatment
XX of neurodegenerative diseases, since they are strongly expressed
XX in the cerebral cortex, hippocampus, striatum and cerebellum.
XX The ligands are also useful as a control or standard in assays,
XX for generation of antibodies, as a mol. wt. marker, for growth
XX in vitro of Htk-receptor positive cells, as research agent,
XX in screening, etc.
XX Sequence 336 AA;
SO Query Match 34.5%; Score 637.5; DB 17; Length 336;
Best Local Similarity 42.1%; Pred. No. 4,66-44;
Matches 139; Conservative 49; Mismatches 129; Indels 13; Gaps 5;
QY 14 GALLLGLVGLGVSLSEPPYWNSSANKRFQAEFGVVLYPOLIGDRDLTLCPRARPSPHSS 73
DB 17 GLMLWLCFTASRSSTVEPTFYWNSSNKFLEPGQGLVLPQIGDKDDITCPKV---DSKTV 73
QY 74 PNYEYKYLVYGAQGRCEAPPAVNLLTCDRPLDLRTTIKQEPYSNPINWGHEFRSHH 133
DB 74 GQYEYKYVMVDKQADRCTIKENNTPLNCARDPDQVKFTIRKEFESFNLMWGEFOKNK 133
QY 134 DYIIANTSDGTREGIESIQGVCLTRGMKVLRLRGQ--SPSGAVPKRPVSEMPEER-DR 190
DB 134 DYYIISTNGSILEGDNGEGVCQTFRANKILMKVGQDASSAGSANHGPTRRPELEACTN 193

```

```

QY 191 GAHSLERKENTLPGDPTSNATSRGAEGLPPSPMAVAGAAAGLALLLGVAGAGAMC 250
DB 194 GRSTTPEFVKNPFGSSITDGNASHGNNLLGSEVALPAGIAGCCTIFVITITLVLL 253
QY 251 WRARRAPSESRHHPGSGFGRGSLGCGGCGMPREAREBELGIALRGGAADPPFCPH 310
DB 254 KYRRHRKHSPQHTTTLSTLSTLATPKRGNN---NGSEPSDVIIPLR---TDSVFCPH 306
QY 311 YEKSGDYGHPVTVQDPGPPSPNNITYKV 340
DB 307 YEKSGDYGHPVTVQEMPSPANNITYKV 336

RESULT 9
AAR55059 standard; Protein; 346 AA.
XX
AC AAR55059;
XX
DT 28-JAN-1995 (first entry)
XX
DE Elk tyrosine kinase receptor ligand.
XX
KW Vectors; elk-L protein; elk; ligands; cell growth; differentiation.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..24 /note="signal peptide"
FT Protein 25..346 /note="mature elk-L protein"
XX
PN M09411384-A.
XX
PD 26-MAY-1994.
XX
PF 15-NOV-1993; 93MO-US10955.
XX
PR 13-NOV-1992; 92US-0977693.
XX
PA (IMMV) IMMUNEX CORP.
XX
PI Baum PR, Beckmann MP, Lyman S;
XX
DR MPI; 1994-183415/22.
XX
DR N-PSDB; AAG65486.
XX
PT New DNA encoding ligand for elk tyrosine kinase receptor - also
PT related polypeptides, vectors, antibodies and probes, useful e.g.
PT in studying cell differentiation or growth
XX
PS Claim 7; Page 30; 35pp; English.
XX
CC The sequence is that of the elk-L protein able to bind elk, a
CC tyrosine kinase receptor. The DNA may be incorporated into vectors
CC which can be used to study the role of elk and its ligands in cell
CC growth and differentiation.
XX
SQ Sequence 346 AA;

Query Match 34.2%; Score 632; DB 15; Length 346;
Best Local Similarity 39.5%; Pred. No. 1.3e-43;
Matches 145; Conservative 48; Mismatches 116; Indels 58; Gaps 9;
QY 8 PGGRVAGALLLGLVLSGL-----SLRPVWNSANKRFQAGGYVYVPOIGRLDIL 61
DB 4 PGGRWLGKMLVAVVWMLCRLATPLAKNLEFVGSNNPKFLSGKGLVYIPKIGDLDIT 63
QY 62 CPDAPRPGEHSSPNVEFYKLYLVGAQGRRCAPPAVNLITCDRPDLRLFTIKQFYS 121
DB 64 CPDAPRAEGRP-----YEVYKLYLVRFQQAACSTVLDPNVLVTCNRPDEIRFTIKQFYS 118

```

```

QY 122 PVLNGEPRSHHDYITITSDGTREGLSLGGVCLITGKAVLLRVGSGPRGAVPRKV 181
DB 119 PNYMLEFRKHHDDYITITSTNSGLEGRNREGVCRITMTKIIMKVGDPPAVVPEOLTT 178
QY 182 SEMPERRGAAHSLR-GEKENTLPGDPTSNATSRGAEGLPPSPMAVAGAAAGLA---- 236
DB 179 SPSSEADANTYVMAAQAGSRGSLGDSGKHETVNNQEKSP-----GASGSSGDPD 231
QY 237 -----LILLGVAGAGA-----MWRRRAPSESRHHPGSGFGRGSLGL 277
DB 232 GFENSKVALFAVAGACVIFLLIIFLTVLLIKLRKRRKHTQO-----RAAALSL 282
QY 278 ---GGGCGMPREAREBELGIALRGGAADPPCPHEKYSGRVHPVTVQDPGPPSP 333
DB 283 STLAPKGGSGTAGTEPBDITIPLR---TTENNYCPHYEKVSGDYGHPVTVQEMPPOSP 339
QY 334 PNITYKV 340
DB 340 ANITYKV 346

RESULT 10
AAR91930 standard; Protein; 346 AA.
XX
AC AAR91930;
XX
DT 11-DEC-1996 (first entry)
XX
DE Human cytokine elk-L ligand (elk-L).
XX
KW Human; cytokine; elk-L ligand; elk-L; tyrosine kinase receptor;
KW neurotrophic; neuroprotective; placenta; radiolabelled probe;
KW treatment; neural tissue; excitotoxicity; injury; disorder;
KW neural culture reagent; immunogenic fragment; antibody.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..24 /label= sig_peptide
FT Peptide 25..346 /label= mat_peptide
XX
PN US5512457-A.
XX
PD 30-APR-1996.
XX
PF 13-NOV-1992; 92US-0977693.
XX
PR 15-MAR-1994; 94US-0213403.
PR 13-NOV-1992; 92US-0977693.
XX
PA (IMMV) IMMUNEX CORP.
XX
PI Baum PR, Beckmann MP, Carpenter MK, Lyman S;
XX
DR MPI; 1996-229866/23.
XX
DR N-PSDB; AAT28770.
XX
PT DNA coding for neurotrophic human elk ligand cytokine - useful as
PT probe to isolate other elk ligand sequences
XX
PS Claim 1; Columns 27-30; 18pp; English.
XX
CC The present sequence is the human cytokine elk-L ligand (elk-L),
CC which binds a member of the tyrosine kinase receptor family. Elk-L
CC exhibits neurotrophic and neuroprotective properties, and has a
CC calculated mol. wt. 35180 daltons and a pI of 9.006. The elk-L
CC cDNA, isolated from a human placental cDNA library, can be
CC radiolabelled and used as a probe for isolating other mammalian
CC elk-L cDNA. Elk-L can be used to treat neural tissue disorders,

```

CC partic. excito-toxicity associated injuries or disorders, and as a  
 CC neutral culture reagent, while immunogenic fragments of elk-L can be  
 CC used to generate specific anti-elk-L antibodies.

CC Sequence 346 AA;

Query Match 34.2%; Score 632; DB 17; Length 346;  
 Best Local Similarity 39.5%; Pred. No. 1.3e-43;  
 Matches 145; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

QY 8 PGGVRVALLILGLVGLVGL-----SLSPYVNSANKRFOAGGYLYLPQIGRLDL 61  
 DB 4 PGGRWLGKMLVAVVWVALCRLATPLAKNLEPVSWSISLRFSGKGLVYIPKIGDLDI 63  
 QY 62 CPRARPPGHSBPNVEFYKLYLVGAQGRCEAPAPNLLTCDRDLRFTIKFOEYS 121  
 DB 64 CPRAEAGRP-----VEYKLYLVREPOAAACSTVLDPNVLVTCNRPEDQIRFTIKFOEFS 118  
 QY 122 PNLWGEPRSHHDYIATSDGTREGLESLOGVCLTSGMKLLRVGQSPRGAAVPRKPV 181  
 DB 119 PNYMGLFEFKKHHDYITSTNSGLEENREGVCRTKIMKVGQDPNAVTPQDLTT 178  
 QY 182 SEMPMERDRGAASLE-PGKENVLPDPTSNATSRGAEGLPPSPNPAVAGAAGLA---- 236  
 DB 179 SRPSKADNTVMATQAPSGRSLGSDGKHETVNOEESGP-----GASGSSGDPD 231  
 QY 237 -----LILLGVAGAGA-----MCMRRRAKPSBSRHPGSGRGSGL 277  
 DB 232 GFNSKVALLFAVAGACVIFLLIIFLVLLKLRKHKHQO-----RAAALSL 282  
 QY 278 -----GGGGGMPREAPRGELGIALRGGAADPPCPHYEKVSGDYHPVYIVODGPPSP 333  
 DB 283 STLASPRGSGGTAGTSPDIIIPLR---TTENNYCPHYEKVSGDYGHVYIVQEMPPQSP 339  
 QY 334 PNITYKV 340  
 DB 340 ANITYKV 346

RESULT 11

ID AAM36055 standard; Protein; 346 AA.

XX AAM36055;  
 AC AAM36055;  
 XX 06-MAR-1998 (first entry)  
 DT Human elk-L protein.  
 XX Human elk-L protein.  
 KW Human, elk-L; cytokine; ligand; tyrosine kinase receptor; fusion protein;  
 KW extracellular domain; immunoglobulin; neurological disease.  
 OS Homo sapiens.  
 XX Key  
 FH Peptide  
 FT 1..24  
 FT /note= "signal peptide"  
 FT 25..346  
 FT /note= "mature protein"  
 FT 25..237  
 FT /note= "extracellular domain; this region is used to  
 generate a fusion protein with the Fc portion  
 of the human immunoglobulin G1"  
 FT Domain  
 FT 238..265  
 FT /note= "transmembrane domain"  
 FT Domain  
 FT 266..346  
 FT /note= "intracellular domain"  
 FT Modified-site  
 FT 139..141  
 FT /note= "Asn is N-glycosylated"  
 FT Cleavage-site  
 FT 266..267  
 FT /note= "KEX2 protease cleavage site"  
 FT Cleavage-site  
 FT 267..268  
 FT /note= "KEX2 protease cleavage site"

FT Cleavage-site 270..271  
 FT /note= "KEX2 protease cleavage site"

XX US5670625-A.

XX 23-SEP-1997.

XX 02-JUN-1995; 95US-0460741.

XX 15-MAR-1994; 94US-0213403.

XX 13-NOV-1992; 92US-0977693.

XX 02-JUN-1995; 95US-0460741.

XX (IMMEX CORP.

XX Baum PR, Beckmann MP, Lyman S;

XX WPI; 1997-479524/44.

XX N-PDB; AAT97976.

XX Soluble fusion proteins of human elk-L ligand and Fc immunoglobulin  
 fragment - and their dimers and oligomers, useful as  
 neuro-protectants and neurotrophic agents

XX Claim 1; Columns 27-30; 18pp; English.

CC This is the amino acid sequence of the human elk-L protein, a new  
 CC cytokine that is the ligand for the elk tyrosine kinase receptor. The  
 CC extracellular domain of the protein (amino acids 1-213) is used to  
 CC generate a fusion protein comprising the Fc polypeptide of the human  
 CC immunoglobulin G1 (extending from the hinge region to the C-terminus).  
 CC The fusion protein (which has the same activities as the natural elk-L  
 CC protein) has neuroprotective and neurotrophic activity so is potentially  
 CC useful for treating a wide range of neurological diseases.

XX Sequence 346 AA;

Query Match 34.2%; Score 632; DB 18; Length 346;  
 Best Local Similarity 39.5%; Pred. No. 1.3e-43;  
 Matches 145; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

QY 8 PGGVRVALLILGLVGLVGL-----SLSPYVNSANKRFOAGGYLYLPQIGRLDL 61  
 DB 4 PGGRWLGKMLVAVVWVALCRLATPLAKNLEPVSWSISLRFSGKGLVYIPKIGDLDI 63  
 QY 62 CPRARPPGHSBPNVEFYKLYLVGAQGRCEAPAPNLLTCDRDLRFTIKFOEYS 121  
 DB 64 CPRAEAGRP-----VEYKLYLVREPOAAACSTVLDPNVLVTCNRPEDQIRFTIKFOEFS 118  
 QY 122 PNLWGEPRSHHDYIATSDGTREGLESLOGVCLTSGMKLLRVGQSPRGAAVPRKPV 181  
 DB 119 PNYMGLFEFKKHHDYITSTNSGLEENREGVCRTKIMKVGQDPNAVTPQDLTT 178  
 QY 182 SEMPMERDRGAASLE-PGKENVLPDPTSNATSRGAEGLPPSPNPAVAGAAGLA---- 236  
 DB 179 SRPSKADNTVMATQAPSGRSLGSDGKHETVNOEESGP-----GASGSSGDPD 231  
 QY 237 -----LILLGVAGAGA-----MCMRRRAKPSBSRHPGSGRGSGL 277  
 DB 232 GFNSKVALLFAVAGACVIFLLIIFLVLLKLRKHKHQO-----RAAALSL 282  
 QY 278 -----GGGGGMPREAPRGELGIALRGGAADPPCPHYEKVSGDYHPVYIVODGPPSP 333  
 DB 283 STLASPRGSGGTAGTSPDIIIPLR---TTENNYCPHYEKVSGDYGHVYIVQEMPPQSP 339  
 QY 334 PNITYKV 340  
 DB 340 ANITYKV 346

RESULT 12

ID AAM19249 standard; Protein; 346 AA.

```

XX AC AAW19249;
XX 18-AUG-1997 (first entry)
XX DT
XX XX Human elk ligand protien.
XX XX
XX XX Human; elk; ligand; elk-L; cytokine; testing; measuring;
XX KM purification; neuroprotection; treatment; diabetic; hereditary;
XX KM nutritional; neuropathy; neurodegenerative disease;
XX XX tissue culture.
XX OS Homo sapiens.
XX XX
XX XX Key Location/Qualifiers
XX FT Peptide 1..24
XX FT /label= sig_peptide
XX FT Peptide 25..346
XX FT /label= mat_peptide
XX FT
XX XX US5627267-A.
XX XX
XX XX 06-MAY-1997.
XX XX
XX PF 13-NOV-1992; 92US-0977693.
XX XX
XX XX 15-MAR-1994; 94US-0213403.
XX PR 13-NOV-1992; 92US-0977693.
XX PR 01-JUN-1995; 95US-0458077.
XX XX
XX PA (IMMUNEX CORP.
XX XX
XX XX Baum PR, Beckmann MP, Lyman S;
XX XX
XX DR WPI; 1997-271366/24.
XX DR N-PSDB; AAT69766.
XX XX
XX FT Human elk ligand protein - for diagnostic or therapeutic use, e.g.
XX PT as neuro-protective agent
XX XX
XX PS Claim 1; Columns 29-32; 18pp; English.
XX XX
XX CC The present sequence is a human elk ligand (elk-L) protein,
XX CC which binds elk, has a calculated molecular weight of 35180 and an
XX CC isoelectric point of 9.006. Elk-L is a cytokine that can be used to
XX CC test cells for elk expression, measure the biological activity of
XX CC elk, purify elk by affinity chromatography and as a neuroprotective
XX CC agent to treat diabetic, hereditary and nutritional neuropathies
XX CC and neurodegenerative diseases. It may also be added to tissue
XX CC cultures to prolong the life of neurons. The elk-L cDNA was
XX CC isolated from a human placental cDNA library, and is present as a
XX CC cDNA insert in the recombinant vector deposited in strain
XX CC ATCC 69085.
XX XX
XX SQ Sequence 346 AA;

```

```

Query Match 34.2%; Score 632; DB 18; Length 346;
Best Local Similarity 39.5%; Pred. No. 1.3e-43;
Matches 145; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

```

```

QY 8 PGGRVAGALLLVGLVSGL-----SLPYWNSANKRFOLEGVLYPOIGRDLTL 61
DB 4 PGGRVAGALLLVAVVWMLCRLATPLAKNLEFVSMNSINPFLSGKLVIYPIKIGDLDIT 63
QY 62 CPRARPPGPHSSPNYEFYKYLIVGAGRCRCPAPANNLLITCDPDLIRFTIKFOEYS 121
DB 64 CPRAEAGRP-----YEFYKYLIVRPEQAAACSTVLDPMVLTICNRPEOIRFTIKFOEYS 118
QY 122 PNLMGHEFRSGHHDYIATSDGTREGESLGGVCLTRGMKVLRYGQSPRGAVPRXRV 181
DB 119 PNIVGLEFFKCHHDYIYISNGSLGLENREGVCRTRIKITMKVGQDBNAVTPQQLTT 178
QY 182 SEMFERDGAASHLF-PGENLPQDPTSNATSGAEGPLPPSPMBAVAGAAAGLA----- 236

```

```

DB 179 GRPSKEADNTVMATQAGSGSLGSDGKHETVVOEKEGP-----GASGSSGSPD 231
QY 237 -----LILLVAGAGAA-----MCMRRRAKPSERHPGSGFRGSGISGL 277
DB 232 GFPSKVALFAVAGACVIFLLIIFLTVLLIKRKRRKHTQQ-----PAAALSL 282
QY 278 -----GGGGMGPRERAPDELIGIALRGGAADPPCPHYEKVSGYGHPIYIVODGPPSP 333
DB 283 SLLAPKGGSGTACGTPEDITIPLR---TENNVCPIREKVSQDIGHPIYIVODGPPSP 339
QY 334 PNIIYKV 340
DB 340 ANIYKV 346

```

```

RESULT 13
AAW4323
ID AAW4323 standard; Protein; 346 AA.
XX AC AAW4323;
XX XX
XX DT 27-MAY-1998 (first entry)
XX XX
XX XX Human elk-L.
XX XX
XX KM Human; elk-L; elk ligand; cytokine; antibody; cell surface receptor;
XX KM tyrosine kinase receptor; neural disease; trophic mechanism.
XX XX
XX OS Homo sapiens.
XX XX
XX XX Key Location/Qualifiers
XX FT Peptide 1..24
XX FT /label= signal
XX FT Protein 25..346
XX FT /label= elk-L
XX FT
XX XX US5728813-A.
XX XX
XX XX 17-MAR-1998.
XX XX
XX PF 12-NOV-1996; 96US-0747240.
XX XX
XX XX 15-MAR-1994; 94US-0213403.
XX PR 13-NOV-1992; 92US-0977693.
XX PR 02-JUN-1995; 95US-0460741.
XX PR 12-NOV-1996; 96US-0747240.
XX XX
XX PA (IMMUNEX CORP.
XX XX
XX XX Baum PR, Beckmann MP, Lyman S;
XX XX
XX DR WPI; 1998-206621/18.
XX DR N-PSDB; AAV15226.
XX XX
XX FT Antibodies to elk ligand polypeptides - bind to tyrosine kinase
XX PT receptors; useful for treating neural disease
XX XX
XX PS Claim 1; Column 31-32; 17pp; English.
XX XX
XX CC The present sequence represents human elk-L (elk ligand). The present
XX CC invention describes antibodies to elk-L polypeptides. The elk-L protein
XX CC is known to be neuroprotective exhibiting its effect through a trophic
XX CC mechanism. Examples of diseases that may be treated with elk-L include
XX CC neuropathies e.g. diabetic, hereditary and nutritional neuropathies,
XX CC neurodegenerative diseases and other diseases characterised by loss of
XX CC function or degeneration of neurons.
XX XX
XX SQ Sequence 346 AA;

```

```

Query Match 34.2%; Score 632; DB 19; Length 346;
Best Local Similarity 39.5%; Pred. No. 1.3e-43;
Matches 145; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

```

```

QY      8  PGGVNGLALLGLVGLVSGT-----SLRPYVNSANKRFOAGGYLYLPQIGRLDLL 61
DB      4  PGRWLGKMLVAMVVALCRLATPLAKNLEPVSWSSLNPKFLSGGLVLYPPKIGKLDIT 63
QY      62  CPAPRPPGPHSSPNVEFYKLYLVGAGRCRCPAPNLLTCDRPLDLRFTTKFQES 121
DB      64  CRRAEGRP-----YEVYKLYLVREPGAACSTVLDPNVLVTCNRPGEIIRFTIKFQFS 118
QY      122  PNLWGHFESHHDYIYIANSDGTREGLESLGGVCLTRGMKYLRLVGGSPRGAVPRKPV 181
DB      119  PVMGIEFFKHHHDYITSTNSGLEENREGVCRTRTMKIMKYGODPNAVTPEDQLTT 178
QY      182  SEMPMERDGAHSLR-PGKENVLPDPPTSNATSGAEGPLPPSPNPAVAGAAGLA----- 236
DB      179  SRPSKADVTVMATQAPGSRGSLDSDGKHETVNOEKSQP-----GASGSSGSDPD 231
QY      237  -----LILLGVAGAGA-----MCMRRRRAPKPSRHPGSGFRGSGSLGL 277
DB      232  GFNSKVALFAAVGAGCVIFLLIIFLVLLKLRKRKHQO-----RAALSL 282
QY      278  -----GGGGMGPREAPPSGLGIALRGGAADPPCPHYEKVSGDHPVYIVQDGPSP 333
DB      283  STLASKGSSGTAGTEPSDIIIPLR--TTENNYCPHYEKVSGDHPVYIVQEMPQSP 339
QY      334  PNIYKCV 340
DB      340  ANIYKCV 346

RESULT 14
AAR82606
ID      AAR82606 standard; Protein; 346 AA.
XX      AC      AAR82606;
XX      DT      16-MAY-1996 (first entry)
XX      DE      Eph transmembrane tyrosine kinase family ligand, Efl-3.
XX      KM      Efl-2; EHK1-L; Eph transmembrane tyrosine kinase family ligand;
XX      KM      neurological disorder; identification; diagnosis.
XX      OS      Homo sapiens.
XX      FH      Key
XX      FT      Peptide
XX      FT      Domain
XX      FT      Misc-difference
XX      FT      /label= "borders main conserved regions"
XX      PN      WO9527060-A2.
XX      PD      12-OCT-1995.
XX      PF      04-APR-1995; 95WO-US04208.
XX      PR      21-OCT-1994; 94US-0327423.
XX      PR      04-APR-1994; 94US-0222075.
XX      PR      12-APR-1994; 94US-0229402.
XX      PR      01-SEP-1994; 94US-0299567.
XX      PA      (REGG-) REGENERON PHARM INC.
XX      PI      Aldrich TM, Davis S, Gale N, Goldfarb M, Maisompierre PC;
XX      PI      Yanoopoulos GP;
XX      DR      WPI; 1995-358635/46.
XX      PT      Ligands which bind Eph family receptors - used in the diagnosis of
XX      PT      neurological disorders

```

```

XX      PS      Disclosure; Fig 2; 56pp; English.
XX      CC      Efl-3 (also known as EHK-L) is an Eph transmembrane tyrosine kinase
XX      CC      family ligand. It has homology with Bel (Efl-1) and Efl-2. Efl-3
XX      CC      appears to comprise a conventional transmembrane protein with a
XX      CC      cytoplasmic domain. Efl-3 is useful for identifying other ligands
XX      CC      for EHK-1, -2, -3. EHK and EHK receptors. The ligands are useful in
XX      CC      promoting a differential function and/or influencing the phenotype,
XX      CC      such as growth and/or proliferation, of receptor bearing cells. They
XX      CC      may be used in the diagnosis, and treatment of neurological disorders.
XX      SO      Sequence 346 AA:
XX      Query Match 34.1%; Score 631; DB 16; Length 346;
XX      Best Local Similarity 40.9%; Pred. No. 1.6e-43;
XX      Matches 139; Conservative 44; Mismatches 105; Indels 52; Gaps 8;
QY      29  SLRPVYVNSANKRFOAGGYLYLPQIGRLDLLCPAPRPPGPHSSPNVEFYKLYLVGAQ 88
DB      31  NLEPVSWSSLNPKFLSGGLVLYPKIGKLDICPRAEGRP-----YEVYKLYLVREPG 85
QY      89  GRGCEAPAPNLLTCDRPLDLRFTTKFQESPNLWGHFESHHDYIYIANSDGTREG 148
DB      86  AACSTVLDENVLAVTCNRPGEIRFTIKFQESPNMGLEFFKHHHDYITSTNSGLE 145
QY      149  ESLGGVCLTRGMKYLRLVGGSPRGAVPRKPVSEMPMERDGAHSLR-PGKENVLPD 207
DB      146  ENREGVCRTRTMKIMKYGODPNAVTPEDQLTTSRPSKADNVKMATQAPGSRGSLGS 205
QY      208  TSNATSGAEGPLPPSPNPAVAGAAGLA-----LILLGVAGAGA----- 248
DB      206  DGKHETVNOEKSQP-----GASGSSGSDPDGFNSKVALFAAVGAGCVIFLLIIFL 258
QY      249  -----MCMRRRRAPKPSRHPGSGFRGSGSLGL-----GGGGMGPREAPPSGLGIALRG 300
DB      259  TVLLKLRKRKHQO-----RAALSLSTLASKGSSGTAGTEPSDIIIPLR-- 307
QY      301  GAADPPCPHYEKVSGDHPVYIVQDGPSPPNYKCV 340
DB      308  -TTENNYCPHYEKVSGDHPVYIVQEMPQSPANIYKCV 346

RESULT 15
AAM00288
ID      AAM00288 standard; Protein; 331 AA.
XX      AC      AAM00288;
XX      DT      19-JAN-1997 (first entry)
XX      DE      Chicken Eph receptor ligand Efl-2.
XX      KM      Eph receptor; ligand; Efl-2; tyrosine kinase; signal transduction;
XX      KM      organogenesis; oncogenesis; tumour; neurological disorder;
XX      KM      diagnosis; gene therapy.
XX      OS      Gallus sp.
XX      PN      WO9626958-A2.
XX      PD      06-SEP-1996.
XX      PF      23-FEB-1996; 96WO-US02673.
XX      PR      27-FEB-1995; 95US-0395415.
XX      PA      (HARD ) HARVARD COLLEGE.
XX      PI      Bergemann AD, Flanagan JG;
XX      PI      WPI; 1996-433391/43.
XX      DR      N-PSDB; AAT40231.

```

XX Bph receptor ligand, ERF-2, and DNA encoding it - used to treat or  
PT prevent neurological diseases, and to modulate binding of ERF-2 to  
PT Bph receptor, e.g. to prevent or treat tumour formation.  
XX  
PS Claim 6, Fig 2A-D; 50pp; English.  
XX  
CC Chicken Bph receptor ligand ERF-2 (AAW00288) is a ligand important  
CC in cellular communication during pattern formation. It is strongly  
CC expressed in the anterior hindbrain and newly-forming somites of  
CC embryos at the early organogenesis stage of development. Its amino  
CC acid sequence was deduced from a cDNA clone (AA740231) isolated from  
CC a chicken cDNA library. The ERF-2 ligands can be used to alter  
CC neurological development, oncogenesis and growth regulation, to  
CC modulate binding of ERF-2 to the Bph receptor, and in diagnostic  
CC assays.  
XX  
SQ Sequence 331 AA;  
XX  
Query Match 34.1%; Score 630.5; DB 17; Length 331;  
Best Local Similarity 40.5%; Pred. No. 1.7e-43;  
Matches 135; Conservative 53; Mismatches 124; Indels 21; Gaps 6;  
XX  
QY 14 GALLIGVGLVSGLSLEPYNNSANKRFOAEGGYLYPQIGRDLDCPPRAPPGPHSS 73  
DB 14 GALTVMRTALAKSIVLDPIYMNNSNPKFLPGQGLVLYPQIGDKDIIICPKV---DSKTA 70  
XX  
QY 74 PNTEFYKLYIVGAQGRCEAPAPNLLITCDRPDLRFPTIKFOEYSPLMWGHEFRSH 133  
DB 71 GUYEYKVVYWDKQADSCAIRKDNTPILNCAKPDQDVFTIKFOEFSNLMGLEFQNK 130  
XX  
QY 134 DYYIATSDGTREGLSLQGVCLTRGMKYLRLVGQSPRGAVPRKPVSEMMER----- 188  
DB 131 DYVYISTNSGLBGLNNOEGGVCTKMKILMKVGDPSAGLPR---SSDPTKPEQEA 187  
XX  
QY 189 -DGAHSLSEPGKENLPDPTSNATSRGAGPLPPSPMAVAGAGLALLLGVAGAGG 247  
DB 188 GTNGKSTSTSPFYKDHSGSSTDG--SKAGHSISLGSEVALFAGIASGCIIFVIITLVV 245  
XX  
QY 248 AMCMRRRAKPSRSRHPGSPGSGSLGIGGGGMPREABEGELGIALRGGAADPPF 307  
DB 246 LLKRYRRHAKSPQHTTTLSTLSTLATPKRSGNN---NGSEPSDIIITPLR---TADSVF 298  
XX  
QY 308 CPHYEKVSGDYGHPIYIVODGPSPPNITYKY 340  
DB 299 CPHYEKVSGDYGHPIYIVODGPSPPNITYKY 331  
XX  
RESULT 16  
AAW06337  
ID AAW06337 standard; Protein; 333 AA.  
XX  
AC AAW06337;  
XX  
DT 17-JAN-1997 (first entry)  
XX  
DE Full length ligand for receptor-type tyrosine kinase protein.  
XX  
KW Receptor-type tyrosine kinase; ligand; coomassie staining; PAS staining;  
XX human.  
XX  
OS Homo sapiens.  
XX  
XX  
XX Key Location/Qualifiers  
XX FT Peptide 1..25  
XX FT /note= "signal peptide"  
XX FT 26..333  
XX FT Protein /note= "receptor-type tyrosine kinase binding protein"  
XX  
XX JP08188596-A.  
XX  
XX 23-JUL-1996.  
XX

PF 13-JAN-1995; 95JP-0003677.  
XX  
XX 09-NOV-1994; 94JP-0275411.  
XX  
XX 19-OCT-1994; 94JP-0253848.  
XX  
XX (ASAH ) ASAH KASEI KOGYO KK.  
XX  
XX WPI: 1996-398601/39.  
XX  
XX N-PSDB; ANT42594.  
XX  
XX  
XX New ligand for receptor type tyrosine kinase - has mol.wt. 22-25  
XX kDa; dalton(s) and is positive for Coomassie and PAS staining  
XX  
XX Example 11; Page 44-46; 51pp; Japanese.  
XX  
XX This sequence represents the full length receptor-type tyrosine kinase  
XX protein binding ligand of the invention (also see AAW06333 and  
XX AAW06334). The mature protein contained within this sequence contains the  
XX N-terminal fragment represented by AAW06332. The proteins of the  
XX invention have a molecular weight of 23500 (plus or minus 1500) Da, and  
XX are positive for Coomassie staining and PAS staining. The proteins of  
XX the invention bind to the fragment of this protein sequence represented  
XX by AAW06330. The proteins of the invention are new ligands of  
XX receptor-type tyrosine kinases, and can be prepared by standard  
XX recombinant techniques.  
XX  
SQ Sequence 333 AA;  
XX  
Query Match 34.0%; Score 629.5; DB 17; Length 333;  
Best Local Similarity 40.9%; Pred. No. 2.1e-43;  
Matches 135; Conservative 52; Mismatches 130; Indels 13; Gaps 5;  
XX  
QY 14 GALLIGVGLVSGLSLEPYNNSANKRFOAEGGYLYPQIGRDLDCPPRAPPGPHSS 73  
DB 14 GVTMLCRTAISKISIVLEPIYMNNSNPKFLPGQGLVLYPQIGDKDIIICPKV---DSKIV 70  
XX  
QY 74 PNTEFYKLYIVGAQGRCEAPAPNLLITCDRPDLRFPTIKFOEYSPLMWGHEFRSH 133  
DB 71 GUYEYKVVYWDKQADSCAIRKDNTPILNCAKPDQDVFTIKFOEFSNLMGLEFQNK 130  
XX  
QY 134 DYYIATSDGTREGLSLQGVCLTRGMKYLRLVGQSPRGAVPRKPVSEMMER-DR 130  
DB 131 DYYIISTNSGLBGLNNOEGGVCTKMKILMKVGDPSAGLPR---SSDPTKPEQEA 129  
XX  
QY 191 GAHSLSEPGKENLPDPTSNATSRGAGPLPPSPMAVAGAGLALLLGVAGAGG 250  
DB 191 GRSTSTSPFYKPNPGSSTDGNAGHSNNILGSEVALFAGIASGCIIFVIITLVVLL 250  
XX  
QY 251 WRRRAKPSRSRHPGSPGSGSLGIGGGGMPREABEGELGIALRGGAADPPF 310  
DB 251 KYRRRAKPSRSRHPGSPGSGSLGIGGGGMPREABEGELGIALRGGAADPPF 303  
XX  
QY 311 YEKVSGDYGHPIYIVODGPSPPNITYKY 340  
DB 304 YEKVSGDYGHPIYIVODGPSPPNITYKY 333  
XX  
RESULT 17  
AAR94655  
ID AAR94655 standard; Protein; 333 AA.  
XX  
AC AAR94655;  
XX  
DT 18-OCT-1996 (first entry)  
XX  
DE Ligand for receptor type tyrosine kinase (TK).  
XX  
XX Receptor type tyrosine kinase; TK; ligand; differentiation;  
XX haematopoietic stem cell; tyrosine; bone marrow; leukaemia.  
XX  
XX Homo sapiens.  
XX  
XX OS  
XX  
XX MO9611212-AI.  
XX

```

XX 18-APR-1996.
XX
XX PF 09-OCT-1995; 95WO-JP02069.
XX
XX PR 22-DEC-1994; 94JP-0320712.
XX PR 07-OCT-1994; 94JP-0244433.
XX PR 26-OCT-1994; 94JP-0262882.
XX
XX (ASAH ) ASAH KASEI KOGYO KK.
XX
XX Ohno M, Sakano S;
XX
XX WPI; 1996-209809/21.
XX DR N-PSDB; AAT18395.
XX
XX PT Ligand peptide binding to receptor-type tyrosine kinase - enhances
XX PT intracellular tyrosine phosphorylation, useful for investigation of
XX PT undifferentiated blood cell behaviour
XX
XX PS Claim 6; Page 163-164; 193pp; Japanese.
XX
XX CC A ligand polypeptide which binds to the extracellular part of a
XX CC specific receptor-type tyrosine kinase and induces phosphorylation
XX CC of tyrosine within the cell can be used in the study of the
XX CC differentiation of blood cells such as the haematopoietic stem
XX CC cells; of disease processes such as leukaemia, and of the biology of
XX CC bone marrow transplantation. The ligand plays a role in the
XX CC differentiation process and the specific ligand target is expressed
XX CC in undifferentiated blood cells.
XX
XX SQ Sequence 333 AA;

Query Match 34.0%; Score 629.5; DB 17; Length 333;
Best Local Similarity 40.9%; Pred. No. 2.1e-43;
Matches 135; Conservative 52; Mismatches 130; Indels 13; Gaps 5;

QY 14 GALLLGLVGLVSGLSLEPYVNSANKRFOAGGYLYPQIGDRDLDCPRARPPGPHSS 73
DB 14 GVLMLVLCRTAISKSIYLEPIYWNSSNKFPLPGGLVLYPQIGDKLDICPKV---DSKTV 70
QY 74 PNYEFKLYVYGAGRGRCRAPRNLLTCDRPLDLAFTTKFOEYSPNLMGHEFRSH 133
DB 71 GQYEYKVMVWDKQADRCTIKKENTPLNCAKPPDDIFTKFOEFSNLMGLEFOQNK 130
QY 134 DYYIIATSDGTREGLESLOGVCLTRGMKVLRLVGQ--SPRGAVPRKPVSEMPMER-DR 190
DB 131 DYYIIISTNGSLGLELDNQGVCQTRAMKILMKVGQDASSAGSTNKKDPTRRPELEAGTN 190
QY 191 GAHSLPEPKENLPDPTSNATSRGAEGLPPPSMPAVAGAAGLALLLVAGAGGCMC 250
DB 191 GRSSTTSPVKPNPGSSTGNSAGHSGNNILSGEVALFAGIASGCIIFVITITLVLL 250
QY 251 WRRRAKPSBSRHPGSGFRGSGSLGGGGGMPREAPGELGIALRGGAADPPFCFH 310
DB 251 KYRRRRKHSPOHTTTLSTLATPKRSGNN---NGSEPSDIIIPLR---TADSVFCFH 303
QY 311 YEKVSGDYGHFVYIVODGPPGSPNNIYKYV 340
DB 304 YEKVSGDYGHFVYIVODGPPGSPNNIYKYV 333

RESULT 18
AAR92743
ID AAR92743 standard; Protein; 333 AA.
XX
XX AAR92743;
XX
XX 21-MAY-1996 (first entry)
XX
XX Human hepatoma transmembrane kinase receptor ligand.
XX
XX Hepatoma transmembrane kinase; Htk; receptor; ligand;

```

```

XX tyrosine kinase; neurodegenerative disease.
XX
XX OS Homo sapiens.
XX
XX PN MO9602645-A2.
XX
XX PD 01-FEB-1996.
XX
XX PF 14-JUL-1995; 95WO-US08812.
XX
XX PR 20-JUL-1994; 94US-0277722.
XX
XX (GETH ) GENENTECH INC.
XX
XX PI Bennett BD, Matthews W;
XX
XX WPI; 1996-105907/11.
XX DR N-PSDB; AAT16471.
XX
XX PT Ligand for the hepatoma trans-membrane kinase receptor - useful for
XX PT stimulating and inhibiting cells carrying the receptor, e.g. for
XX PT treating neuro-degenerative disease
XX
XX PS Claim 6; Fig 2(A-B); 88pp; English.
XX
XX CC Mouse (AAT16470) and human (AAT16471) Htk ligand which bind to, and
XX CC activate, the Htk receptor, have been identified in a variety of
XX CC tissues using a soluble Htk-Fc fusion protein.
XX CC The predicted mol.wt. of the murine Htk ligand protein following a
XX CC signal peptide cleavage is 34 kD with an estimated pI of 8.9.
XX CC The murine and human ligands show 96% homology at the amino acid
XX CC level.
XX CC The DNA is used to produce recombinant ligands; for tissue-
XX CC specific typing (patric. as a marker for breast cancer) and as a
XX CC marker for human chromosome 13. The ligands (patric. in soluble
XX CC form) are used to activate the tyrosine kinase domain of the
XX CC Htk receptor, i.e. to stimulate or inhibit growth, differentiation,
XX CC and/or activation of cells contg. the receptor, e.g. treatment
XX CC of neurodegenerative diseases, since they are strongly expressed
XX CC in the cerebral cortex, hippocampus, striatum and cerebellum.
XX CC The ligands are also useful as a control or standard in assays,
XX CC for generation of antibodies, as a mol. wt. marker, for growth
XX CC in vitro of Htk-receptor positive cells, as research agent,
XX CC in screening, etc.
XX
XX SQ Sequence 333 AA;

Query Match 34.0%; Score 629.5; DB 17; Length 333;
Best Local Similarity 40.9%; Pred. No. 2.1e-43;
Matches 135; Conservative 52; Mismatches 130; Indels 13; Gaps 5;

QY 14 GALLLGLVGLVSGLSLEPYVNSANKRFOAGGYLYPQIGDRDLDCPRARPPGPHSS 73
DB 14 GVLMLVLCRTAISKSIYLEPIYWNSSNKFPLPGGLVLYPQIGDKLDIOPKV---DSKTV 70
QY 74 PNYEFKLYVYGAGRGRCRAPRNLLTCDRPLDLAFTTKFOEYSPNLMGHEFRSH 133
DB 71 GQYEYKVMVWDKQADRCTIKKENTPLNCAKPPDDIFTKFOEFSNLMGLEFOQNK 130
QY 134 DYYIIATSDGTREGLESLOGVCLTRGMKVLRLVGQ--SPRGAVPRKPVSEMPMER-DR 190
DB 131 DYYIIISTNGSLGLELDNQGVCQTRAMKILMKVGQDASSAGSTNKKDPTRRPELEAGTN 190
QY 191 GAHSLPEPKENLPDPTSNATSRGAEGLPPPSMPAVAGAAGLALLLVAGAGGCMC 250
DB 191 GRSSTTSPVKPNPGSSTGNSAGHSGNNILSGEVALFAGIASGCIIFVITITLVLL 250
QY 251 WRRRAKPSBSRHPGSGFRGSGSLGGGGGMPREAPGELGIALRGGAADPPFCFH 310
DB 251 KYRRRRKHSPOHTTTLSTLATPKRSGNN---NGSEPSDIIIPLR---TADSVFCFH 303
QY 311 YEKVSGDYGHFVYIVODGPPGSPNNIYKYV 340

```





```

FT Protein 26..333
ET /label= receptor-type tyrosine kinase
XX
XX JP08280385-A.
XX
XX 29-OCT-1996.
XX
XX 13-APR-1995; 95JP-0087878.
XX
XX 13-APR-1995; 95JP-0087878.
XX
XX (ASAH ) ASAH KASEI KOGYO KK.
XX
XX WPI; 1997-014848/02.
XX
XX N-PSDB; AAT51236.
XX
XX LPM cell culture medium - for culturing cells which produce
PT receptor-type tyrosine kinase ligand
XX
XX Disclousure; Page 40-41; 45pp; Japanese.
XX
XX AA011307 and AA011308 represent receptor-type tyrosine kinases. These
CC sequences are included in the LPM medium of the invention. The medium
CC is used to culture cells capable of producing a ligand for receptor-type
CC tyrosine kinase. The ligand serves as a physiologically active
CC substance to induce enzyme activity and autophosphorylation of tyrosine
CC kinase, which is involved in the development and differentiation of
CC animal and insect cells. The ligand can also be used as in a
CC pharmaceutical preparation for promoting growth of undifferentiated
CC cells. The medium may also include an inducer, such as insulin, an
CC interleukin, a fibroblast growth factor, or interferon-gamma.
XX
XX Sequence 333 AA:
SQ
Query Match 34.0%; Score 629.5; DB 18; Length 333;
Best Local Similarity 40.9%; Pred. No. 2.1e-43;
Matches 135; Conservative 52; Mismatches 130; Indels 13; Gaps 5;
QY 14 GALLLVGLVGLVSLPEVYVNSANKRFOAGGYLVLPQIGDRDLDCPRARPPGPHSS 73
DB 14 GVLVLCRTAISKSYLIERIYVNSNSKFLPGQGLVLPQIGDKDILCPKV---DSKTV 70
QY 74 PNYEFKYLIVGAGQGRCEAPRNLLTCRPLDLRFTTKFOEYSPNLMGHEFRSHH 133
DB 71 GQGEYKYVMVKDQADRTIKKENTPLNCAKPPQDIFTKFOEFSBNLMGLEFOQNK 130
QY 134 DVIYIATSDGTREGLESLOGVCLTRGMKVLRLVQ--SPRGAVPRKPVSEMPYER-DR 190
DB 131 DVIYIISTNGSLGLELDNGCGVCOQTRAMKILMKVGQDASASGSTRNKDPTRRPELFACTN 190
QY 191 GAHSLPEKENVLPDDPTSNATSRGAEGLPPLPPSPAVAGAAAGLALLLGVAGAGGAMC 250
DB 191 GRSSTSPFVKKNPGSGTGNAGSHGNNILGSEVALFAGISGCIIFVIITLVLL 250
QY 251 WRRRRKAPESHPGSGFRGSGSLGLGGGGMGRBAEPGELGALRGGAADPFCPH 310
DB 251 KYRRRRKXSPQHTTLLSLSTATPKRSGNM---NGSEPSDIIIPLR---TADSVFCPH 303
QY 311 YEKVSGDYGHPIYIVODGPPQSPNNIYKYV 340
DB 304 YEKVSGDYGHPIYIVQEMPPQSPANNIYKYV 333

```

RESULT 21  
AA06334

ID AA06334 standard; protein; 308 AA.

XX AA06334;

XX 17-JAN-1997 (first entry)

XX Ligand #2 for receptor-type tyrosine kinase protein.

```

KW Receptor-type tyrosine kinase; ligand; coomassie staining; PAS staining;
KW human.
XX
XX Homo sapiens.
XX
XX JP08188596-A.
XX
XX 23-JUL-1996.
XX
XX 13-JUN-1995; 95JP-0003677.
XX
XX 09-NOV-1994; 94JP-0275411.
XX
XX 19-OCT-1994; 94JP-0253848.
XX
XX (ASAH ) ASAH KASEI KOGYO KK.
XX
XX WPI; 1996-388601/39.
XX
XX New ligand for receptor type tyrosine kinase - has mol.wt. 22-25
PT kilo(dalton(s) and is positive for Coomassie and PAS staining
XX
XX Claim 7; Page 44; 51pp; Japanese.
XX
XX AA06333 and AA06334 represent receptor-type tyrosine kinase receptor
CC binding ligands of the invention. These sequences both have the
CC N-terminal sequence represented by AA06332. The ligands of the
CC invention recognise the fragment of the receptor type kinase receptor
CC represented by AA06330 (see AA06331 for full length sequence). The
CC proteins of the invention have a molecular weight of 23500 (plus or minus
CC 1500) Da, and are positive for Coomassie staining and PAS staining. The
CC protein is a new ligand of receptor-type tyrosine kinases, and can be
CC prepared by standard recombinant techniques.
XX
XX Sequence 308 AA:
SQ
Query Match 33.5%; Score 619.5; DB 17; Length 308;
Best Local Similarity 42.0%; Pred. No. 1.2e-42;
Matches 132; Conservative 48; Mismatches 121; Indels 13; Gaps 5;
QY 30 LEPVYVNSANKRFOAGGYLVLPQIGDRDLDCPRARPPGPHSPNYEFKYLIVGAGAG 89
DB 5 LEPIVYVNSNSKFLPGQGLVLPQIGDKDILCPKV---DSKTVGQYKYVMVKDQ 61
QY 90 RCCEAPRNLLTCRPLDLRFTTKFOEYSPNLMGHEFRSHHDIYIATSDGTREGLE 149
DB 62 DRCITIKKENTPLNCAKPPQDIFTKFOEFSBNLMGLEFOQNKDYIISTNGSLGLELD 121
QY 150 SLOGVCLTRGMKVLRLVQ--SPRGAVPRKPVSEMPYER-DRGAHSLPEKENVLPD 206
DB 122 NQEGVCOQTRAMKILMKVGQDASASGSTRNKDPTRRPELFACTNNGRSTSPVKNPGS 181
QY 207 PTSNATSRGAEGLPPLPPSPAVAGAAAGLALLLGVAGAGGAMCWRRRKAPESHPGE 266
DB 182 STDGNSAGSHGNNILGSEVALFAGISGCIIFVIITLVLLKYRRRRKXSPQHTT 241
QY 267 GSFGRGSLGLGGGGMGRBAEPGELGALRGGAADPFCPHYKVGSDYGHPIYIVQ 326
DB 242 LSLSTATPKRSGNM---NGSEPSDIIIPLR---TADSVFCHEKYZKVGSDYGHPIYIVQ 294
QY 327 DGPPQSPNNIYKYV 340
DB 295 EMPQSPANNIYKYV 308

```

RESULT 22

ID AAR94656 standard; protein; 308 AA.

XX AAR94656;

XX 18-OCT-1996 (first entry)

XX Ligand for receptor type tyrosine kinase (TK) (mature protein).

XX Receptor type tyrosine kinase; TK; ligand; differentiation;  
KM haematopoietic stem cell; tyrosine; bone marrow; leukaemia.  
XX  
OS Homo sapiens.  
XX NO9611212-A1.  
XX 18-APR-1996.  
XX  
XX 09-OCT-1995; 95MO-JP02069.  
XX  
XX 22-DEC-1994; 94JP-0320712.  
PR 07-OCT-1994; 94JP-0244433.  
PR 26-OCT-1994; 94JP-0262882.  
XX  
XX (ASAH ) ASAH KASEI KOGYO KK.  
XX  
XX Ohno M, Sakano S;  
PI  
XX WPI, 1996-209809/21.  
DR N-PSDB; AAT18395.  
XX  
XX Ligand peptide binding to receptor-type tyrosine kinase - enhances  
PT intracellular tyrosine phosphorylation, useful for investigation of  
PT undifferentiated blood cell behaviour  
XX  
XX Disclosure; Page 162-163; 193pp; Japanese.  
XX  
XX A ligand polypeptide which binds to the extracellular part of a  
CC specific receptor-type tyrosine kinase and induces phosphorylation  
CC of tyrosine within the cell can be used in the study of the  
CC differentiation of blood cells such as the haematopoietic stem  
CC cells; of disease processes such as leukaemia, and of the biology of  
CC bone marrow transplantation. The ligand plays a role in the  
CC differentiation process and the specific ligand target is expressed  
CC in undifferentiated blood cells.  
XX  
XX Sequence 308 AA;

Query Match 33.5%; Score 619.5; DB 17; Length 308;  
Best Local Similarity 42.0%; Pred. No. 1.2e-42;  
Matches 132; Conservative 48; Mismatches 121; Indels 13; Gaps 5;

QY 30 LEPYNNANRPFQAGGVLYLPOIGDRLDLCPRARPPPHSHSPYEPKYLYVGAGAG 89  
DB 5 LEPYNNANRPFQAGGVLYLPOIGDRLDLCPRARPPPHSHSPYEPKYLYVGAGAG 61  
QY 90 RCEAPPAPNLLITCDRFDLRTTINQFQYSPNLMGHEFRSHHDYIATSDGTREGLE 149  
DB 62 DRTIKKENTPPLNCAKPDQDIKFTIKFQFSPNLMGHEFRSHHDYIATSDGTREGLE 121  
QY 150 SLOGGVLTGGMKYLAVGQ--SPRGAVPRKRVSEMPMR--DRGAHSLPEKLELPGD 206  
DB 122 NQSGGVQVTAMKILMGVGDASAGSTRNKDPTRRPELGAITNGSSTTSPPVKENPGS 181  
QY 207 PTNATRGAGELPPSPMAVAGAAGLALLLGVAGAGACWMPRRAPKSPSESHPGP 266  
DB 182 STDGNMAGHGNLIGSEVALFAGIASGCTIFVITLVLVLLKTRRRKASPHHTTT 241  
QY 267 GSGRGGSLGAGGGGKGPREADPBGELALRGGAADPPFCHYKRVSGDYGHPIYIQ 326  
DB 242 LLSLSTLTPRSGNN---NGSEPSDIIPLR---TADSVFCHYKRVSGDYGHPIYIQ 294  
QY 327 DGPPSPENIYKYV 340  
DB 295 EMPPOSANIIYKYV 308

RESULT 23  
AA71438  
ID AAY71438 standard; peptide; 89 AA.  
XX

AC AAY71438;  
XX  
XX 04-OCT-2000 (first entry)  
DT  
XX  
DE Human ephrin B3 C-terminal cytoplasmic domain.  
XX  
XX PDZ domain; B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;  
XX cellular process; nerve cell interaction; regeneration of nerve cell;  
KM axonogenesis; antiproliferative; proliferative disorder; treatment;  
XX differentiation disorder; human; cytoplasmic domain.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH Region 57..75  
FT /note= "Conserved region"  
FT Modified-site 60  
FT /note= "Conserved tyrosine phosphorylation site"  
FT Modified-site 67  
FT /note= "Conserved tyrosine phosphorylation site"  
FT Modified-site 72  
FT /note= "Conserved tyrosine phosphorylation site"  
FT Region 78..82  
FT /note= "Conserved region"  
FT Region 84..89  
FT /note= "Conserved region"  
FT Domain 86..89  
FT /label= PDZ\_domain  
FT Modified-site 86  
FT /note= "Conserved tyrosine phosphorylation site"  
FT Modified-site 87  
FT /note= "Conserved tyrosine phosphorylation site"  
XX  
XX W0200031124-A2.  
XX  
XX 02-JUN-2000.  
XX  
XX 19-NOV-1999; 99MO-CA01101.  
XX  
XX 20-NOV-1998; 98US-0109158.  
XX  
XX (MOUN ) MOUNT SINAI HOSPITAL.  
XX  
XX Lin D, Pawson A;  
XX WPI; 2000-400038/34.  
XX  
XX Example; Fig 1; 59pp; English.  
XX  
XX The patent discloses a complex comprising of a B class ephrin and PDZ  
CC domain containing protein. B class ephrine function as ligands for Eph  
CC receptor tyrosine kinases (RTK) and possesses a transmembrane element and a  
CC highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ  
CC binding site. This complex is used in methods to modulate the interaction  
CC of a B class ephrin and PDZ domain containing protein and to identify  
CC modulator compounds. It is also used for modulating cellular processes  
CC like, axonogenesis, nerve cell interactions and regeneration of nerve  
CC cells. The complex is also useful for treating proliferative or  
CC differentiative disorders associated with this protein complex.  
CC The present sequence is the human ephrin B3 C-terminal cytoplasmic  
CC domain, comprising conserved sequences reminiscent of known or predicted  
CC binding sites for PDZ domains. Ephrin B3 is also known as NLERK-2,  
CC ELK-13, EFL-6, ELF-3 and LERK-8.  
XX  
XX Sequence 89 AA;  
XX

Query Match 26.6%; Score 492; DB 21; Length 89;  
Best Local Similarity 97.8%; Pred. No. 7.5e-33;  
Matches 87; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

QY 252 RRRRAKPSRHPGSGFGRGSLGLGGGGGMPREAPFELGIALRGGAADPPFCPHY 311
DB 1 RRRRAKPSRHPGSGFGRGSLGLGGGGGMPREAPFELGIALRGGAADPPFCPHY 60
QY 312 EKVSGDYGHVYIVODGPSPSPNIIYKY 340
DB 61 EKVSGDYGHVYIVODGPSPSPNIIYKY 89

RESULT 24
ID AAY96782 standard; Protein; 658 AA.
AC AAY96782;
DT 26-SEP-2000 (first entry)
DE Ephrin-B2-Ephrin-B2-FC fusion protein.
XX Ephrin-B2; ELK receptor; ligand; dimer; Fc domain; fusion protein;
XX Efl-6 antagonist; neurological.
OS Chimeric - Homo sapiens.
XX Chimeric - Synthetic.
FH Key Location/Qualifiers
FT Peptide 1..30
FT /label= Signal peptide
FT /note= "Derived from ephrin-B2 ectodomain 1"
FT Protein 31..225
FT /label= Ephrin-B2_Ectodomain-1
FT Peptide 226..228
FT /label= Bridging_peptide
FT Protein 229..423
FT /label= Ephrin-B2_ectodomain_2
FT Peptide 424..426
FT /label= Bridging_peptide
FT Region 427..656
FT /label= Human_IgG1_Fc_region

WO200037642-A1.
29-JUN-2000.
XX 23-DEC-1999; 99WO-US30900.
XX 23-DEC-1998; 96US-0113387.
XX (REGE-) REGENERON PHARM INC.
XX Davis SJ, Gale NW, Yancopoulos GD, Stahl N;
XX WPI; 2000-442670/38.
XX N-PSDB; AAA51346.
XX Polynucleotide encoding a fusion polypeptide, useful for promoting
XX differential function and influencing phenotype, comprises two subunits
XX containing at least one copy of the receptor binding domain of a ligand
XX Example 12; Fig 15A-E; 97pp; English.
XX Production of homogenous forms of clustered ligands is broadly applicable
XX to improve the affinity and/or increase the activity of a ligand as
XX compared to the native form of the ligand. Ephrin fusion proteins have
XX been constructed, which may be useful for treating neurological
XX disorders. The ephrin fusion proteins are preferably capable of binding
XX to B1x receptor and are especially Efl-6 antagonists. The fusion proteins
XX were constructed after it was demonstrated that similar improved
XX activities could be achieved using Tie-2 receptor ligands.
XX Angiopoietin-1 (Ang-1) is a Tie-2 receptor ligand and is an agonist for
XX Tie-2, whereas angiopoietin-2 (Ang-2) is a naturally occurring antagonist
XX of the Tie-2 receptor. The fibrinogen domains (FD) of Ang-1 and Ang-2 are
XX the receptor-binding domains and dimerized versions, of e.g. Ang-1-FD-Fc

```

```

CC (Ang-1 fibrinogen domain fused to an Fc domain), can bind to the Tie-2
CC receptor with much higher affinity than monomeric Ang-1-FD (dimerization
CC occurs between the Fc components of adjacent molecules). However, the Tie-2
CC Ang-1-FD is not able to induce phosphorylation (activate) the Tie-2
CC receptor on endothelial cells unless it is further clustered with goat
CC anti-human Fc antibodies. The novel fusion proteins, mutant versions of
CC Ang-1-FD and Ang-2-FD, were designed that were intrinsically more highly
CC clustered. Tie-2 agonist fusion proteins may be used as haematopoietic
CC factors. Tie-2 receptor antagonist fusion proteins may be used to
CC diagnose or treat, e.g. myeloproliferative or other proliferative
CC disorders of blood forming organs, e.g. thrombocythemias, polycythemias
CC and leukemias.
XX
XX Sequence 658 AA;
SQ
Query Match 24.5%; Score 453.5; DB 21; Length 658;
Best Local Similarity 34.6%; Pred. No. 1,2e-28;
Matches 111; Conservative 32; Mismatches 79; Indels 99; Gaps 8;
QY 5 HSGPGGVAVGALLLGLVLVSGLSLEPYVWNSANKRFOABGVLYLPQIGRLDLCPR 64
DB 218 HSGNN-----LLGPGIV---LEPIYVNSNKKFLPGQGLVLYPQIGDKLDICPK 265
QY 65 ARPPGHSSPNVPEFYLYLVGAQGRCEAPAPNILLTCDDPPDDLRFTTFQEXSPYL 124
DB 266 V--DSKTGVQYRYVYVWDKQADRCTIKENTPLNCAAPDDVKTITFOEESPVL 322
QY 125 WGEHFRSHDYYIIATSDTREGLESIGGVCLTRGMKYLIVGQSPRGAVRRVSEH 184
DB 323 WGLEPQKNQDYIIISTNSLSLELDNQEGVQGTAMKTIKLVGQD----- 368
QY 185 EMERDGAHSLERPKENLPGDPTSNATSRGABGLPPPSMPAVVGAAGLALLLVGAV 244
DB 369 -----ASSGASARNHGPFRRELS-----AG 389
QY 245 AGGAMCRRRAKPSRHPGSGFGRGSLGLGGGGGMPREAPFELGIALRGGAAD 304
DB 390 TNG-----RSSITSPFYKPNPQSSITDGNASHGNNLL-----GGPGE 427
QY 305 P-----PFCPHYEKVSG 316
DB 428 PKSCDHTTCPCPAPDELIG 448

RESULT 25
ID AAY96781 standard; Protein; 683 AA.
AC AAY96781;
DT 26-SEP-2000 (first entry)
DE Ephrin-B1-Ephrin-B1-FC fusion protein.
XX Ephrin-B1; ELK receptor; ligand; dimer; Fc domain; fusion protein;
XX Efl-6 antagonist; neurological.
XX Chimeric - Homo sapiens.
XX Chimeric - Synthetic.
FH Key Location/Qualifiers
FT Peptide 1..29
FT /label= Signal peptide
FT /note= "derived from ELK-L Ectodomain 1"
FT Protein 30..237
FT Peptide 238..240
FT /label= Bridging_peptide
FT Protein 241..448
FT /label= ELK-L_Ectodomain_2
FT Peptide 449..451
FT /label= Bridging_peptide
FT Region 452..683

```

FT XX /label= Human\_IgG1\_Fc\_region  
 XX MO200037642-A1.  
 XX 23-JUN-2000.  
 XX 23-DEC-1999; 99MO-US0900.  
 XX 23-DEC-1998; 98US-0113387.  
 XX (REGG-) REGENERON PHARM INC.  
 XX Davis SJ, Gale NW, Yancopoulos GD, Stahl N;  
 XX WPI: 2000-442670/36.  
 XX N-PSDB; AAA51345.  
 XX Polynucleotide encoding a fusion polypeptide, useful for promoting  
 PT differential function and influencing phenotype, comprises two subunits  
 PT containing at least one copy of the receptor binding domain of a ligand  
 XX  
 PS Example 12; Fig 14A-E; 97pp; English.  
 CC Production of homogenous forms of clustered ligands is broadly applicable  
 CC to improve the affinity and/or increase the activity of a ligand as  
 CC compared to the native form of the ligand. Ephrin fusion proteins have  
 CC been constructed, which may be useful for treating neurological  
 CC disorders. The ephrin fusion proteins are preferably capable of binding  
 CC to Etk receptor and are especially Etk-6 antagonists. The fusion proteins  
 CC were constructed after it was demonstrated that similar improved  
 CC activities could be achieved using Tie-2 receptor ligands.  
 CC Angiopoietin-1 (Ang-1) is a Tie-2 receptor ligand and is an agonist for  
 CC of the Tie-2 receptor. The fibrinogen domains (FD) of Ang-1 and Ang-2 are  
 CC the receptor-binding domains and dimerized versions, of e.g. Ang-1-FD-Fc  
 CC (Ang-1 fibrinogen domain fused to an Fc domain), can bind to the Tie-2  
 CC receptor with much higher affinity than monomeric Ang-1-FD (dimerization  
 CC occurs between the Fc components of adjacent molecules). However,  
 CC Ang-1-FD-Fc is not able to induce phosphorylation (activate) the Tie-2  
 CC receptor on endothelial cells unless it is further clustered with goat  
 CC anti-human Fc antibodies. The novel fusion proteins, mutant versions of  
 CC Ang-1-FD and Ang-2-FD, were designed that were intrinsically more highly  
 CC clustered. Tie-2 agonist fusion proteins may be used as haematopoietic  
 CC factors. Tie-2 receptor antagonist fusion proteins may be used to  
 CC diagnose or treat, e.g. myeloproliferative or other proliferative  
 CC disorders of blood forming organs, e.g. thrombocythemias, polycythemias  
 CC and leukemias.  
 CC  
 SQ Sequence 683 AA;  
 Query Match 24.5%; Score 452.5; DB 21; Length 683;  
 Best Local Similarity 44.0%; Pred. No. 1.5e-28;  
 Matches 92; Conservative 32; Mismatches 72; Indels 13; Gaps 3;  
 QY 27 GLSLEPYMNSANKRFOAEGGYLVLPQIGRLDILCPRAPRPPHSSVYEFKYLIVGG 86  
 DB 240 GKULEPWSNLSNPKFSGGLVYPIKIGKLDIICPRLEAGRP-----YEEYKYLIVRP 294  
 QY 87 AGGRCEAPAPNILLTCRPPDLRLFTTIKFOEYSPNLMGHEFRSHDYIATSDGRE 146  
 DB 295 EQAAACSTVADPVIYVTCNRPBGEIRFTIKFOEYSPNLMGHEFRSHDYIATSDGRE 354  
 QY 147 GLSLEGGVCLTGMKYLIRVGGSPRGAVPRKPSVEMPERDRGAASLE-PGKENLPG 205  
 DB 355 GLNREGVCRITRMKIMKVGDPNAVYPEQLTSRPSKEADNTVMAVTAQAPRGSLG 414  
 QY 206 DPTSNATRGAEGLPPSPMPAVAGAAG 234  
 DB 415 DSDGKETVNOEKSQP-----GASGG 436

ID AAE24019 standard; Protein; 229 AA.  
 XX AAE24019;  
 AC 23-SEP-2002 (first entry)  
 DT 23-SEP-2002 (first entry)  
 XX Murine ephrin B2 ligand.  
 DE Murine ephrin B2 ligand.  
 XX Murine; extracellular region; Eph B receptor; ephrin B ligand; tumour;  
 XX ocular neovascularisation; gene therapy; cytostatic.  
 XX Mus musculus.  
 OS Mus musculus.  
 PN WO200226627-A1.  
 XX 04-APR-2002.  
 PD 28-SEP-2001; 2001WO-BP11252.  
 XX 29-SEP-2000; 2000CH-0001910.  
 XX (NOVS ) NOVARTIS AG.  
 PA (NOVS ) NOVARTIS-ERRINDUNGEN VERM GES MBH.  
 XX Martiny-Baron G, Wood JM, Liao G;  
 PI WPI: 2002-405047/43.  
 DR N-PSDB; AAD38768.  
 XX New soluble polypeptides of the extracellular region of the Eph B  
 PT receptor and ephrin B ligands, useful in gene therapy for treating a  
 PT tumour disease or ocular neovascularization in mammals, including humans  
 PT  
 PS Claim 6; Page 48-49; 55pp; English.  
 XX The invention relates to new isolated soluble polypeptides comprising an  
 CC amino acid sequence of an extracellular region of an Eph B receptor or an  
 CC amino acid sequence of an ephrin B ligand, or essentially similar amino  
 CC acid sequences of extracellular region of an Eph B receptor or an amino  
 CC acid sequence of an ephrin B ligand. Extracellular region of an Eph B  
 CC receptor is capable of binding with high affinity to a mammalian ephrin B  
 CC ligand. Ephrin B ligand is capable of binding with high affinity to a  
 CC mammalian Eph B receptor. The nucleic acid encoding the polypeptides or  
 CC the polypeptides are useful for producing pharmaceutical compositions for  
 CC treating a tumour disease or ocular neovascularisation in mammals,  
 CC including humans. The vector is useful in producing a pharmaceutical  
 CC composition for treating the diseases cited above in mammals,  
 CC particularly humans, by means of gene therapy. The present sequence is  
 CC murine ephrin B2 ligand.  
 CC  
 SQ Sequence 229 AA;  
 Query Match 24.2%; Score 447; DB 23; Length 229;  
 Best Local Similarity 36.4%; Pred. No. 1.1e-28;  
 Matches 100; Conservative 33; Mismatches 80; Indels 62; Gaps 4;  
 QY 14 GALLILGLVLSGLSEPYMNSANKRFOAEGGYLVLPQIGRLDILCPRAPRPPHSS 73  
 DB 17 GLVWLCTRAISRIIVLEPYMNSNGKFLPGGLVYPIKIGKLDIICRKV--DSKTV 73  
 QY 74 PNYEFKYLIVGAGQGRCEAPAPNILLTCRPPDLRLFTTIKFOEYSPNLMGHEFRSHH 133  
 DB 74 GQYEVYKVMYDQADRCITKENTPLNLCARDQVKEFTIKFOEYSPNLMGHEFRSHH 133  
 QY 134 DYIATSDGREGLSLOGVCLTGMKYLIRVGGSPRGAVPRKPSVEMPERDRGAAG 193  
 DB 134 DYIATSDGREGLSLOGVCLTGMKYLIRVGGSPRGAVPRKPSVEMPERDRGAAG 193  
 QY 194 HSLPEKENTPGDPTSNATRGAEGLPPSPMPAVAGAAGLALLILGVAGAGAMCWR 253  
 DB 171 -----ASSAGSRNNGPTRPELE-----AGTNG----- 194

QY 254 RRAKPESESRHPPGSGFRGSGSLGGGGGMGPREA 288  
 DB 195 RSSSTSPFVKPNPGSSSTDGNSAGHSNNILGSEVA 229

RESULT 27  
 AAE24020  
 ID AAE24020 standard, Protein, 229 AA.

AC AAE24020;  
 XX  
 DT 23-SEP-2002 (first entry)

DE Human ephrin B2 ligand.

KW Human; extracellular region; Eph B receptor; ephrin B ligand; tumour;  
 KM ocular neovascularisation; gene therapy; cyostatic.

OS Homo sapiens.  
 XX  
 PN WO200226827-A1.

PD 04-APR-2002.

PF 28-SEP-2001; 2001WO-EP11252.

PR 29-SEP-2000; 2000CH-0001910.

PA (NOVS ) NOVARTIS AG.  
 PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.

PI Martiny-baron G, Wood JM, Liau G;

DR WPI; 2002-405647/43.  
 DR N-PsDB; AAD38769.

XX New soluble polypeptides of the extracellular region of the Eph B  
 PT receptor and ephrin B ligands, useful in gene therapy for treating a  
 PT tumor disease or ocular neovascularization in mammals, including humans

PS Claim 6; Page 51-52; 55pp; English.

XX The invention relates to new isolated soluble polypeptides comprising an  
 CC amino acid sequence of an extracellular region of an Eph B receptor or an  
 CC amino acid sequence of an ephrin B ligand, or essentially similar amino  
 CC acid sequences of extracellular region of an Eph B receptor or an amino  
 CC acid sequence of an ephrin B ligand. Extracellular region of an Eph B  
 CC receptor is capable of binding with high affinity to a mammalian ephrin B  
 CC ligand. Ephrin B ligand is capable of binding with high affinity to a  
 CC mammalian Eph B receptor. The nucleic acid encoding the polypeptides or  
 CC the polypeptides are useful for producing pharmaceutical compositions for  
 CC treating a tumor disease or ocular neovascularisation in mammals,  
 CC including humans. The vector is useful in producing a pharmaceutical  
 CC composition for treating the diseases cited above in mammals,  
 CC particularly humans, by means of gene therapy. The present sequence is  
 CC human ephrin B2 ligand.

XX Sequence 229 AA;

Query Match 23.9%; Score 443; DB 23; Length 229;

Best Local Similarity 35.6%; Pred. No. 2.4e-28;

Matches 96; Conservative 36; Mismatches 79; Indels 62; Gaps 4;

QY 14 GALLLLGLVGLVSGSLSEPYVNSANKRFQAEAGVYLPQIGDRLLDLCPRARPSPHSS 73

DB 14 GVLAVLCLRTAISKSIIVLEPIYVNSNKKFLPGQGLVLPQIGDKLIIICPKV--DSKTV 70

QY 74 PNVSEFYLIVYVGAAGRCEAPNPULLCDBPDLDLRTIKFOEYSPYLKMGHERSRSH 133

DB 71 GQYVYVYVWVDKQARCTIKKENTPLNCARPDQDIKTIKFOEYSPYLKMGHERSRSH 130

QY 134 DYLIATSDGTREGLELQGGVCLTRGMKVLLRVGQSPRGAAVPRKPVSEMPERDRGAA 193

DB 131 DYLIATSDGTREGLELQGGVCLTRGMKVLLRVGQSPRGAAVPRKPVSEMPERDRGAA 166  
 QY 194 HSLEPGKENVLPDPTSNATSRGAEGPLPPSPMPAVGAAGLALLILVAGAGAMCMRR 253  
 DB 167 -----DASAGSTRNKDPTRPELE-----AGTNG----- 191

QY 254 RRAKPESESRHPPGSGFRGSGSLGGGGGMGPREA 288  
 DB 192 RSSSTSPFVKPNPGSSSTDGNSAGHSNNILGSEVA 226

RESULT 28  
 AAM06333  
 ID AAM06333 standard, protein, 195 AA.

AC AAM06333;

DT 17-JAN-1997 (first entry)

DE Ligand #1 for receptor-type tyrosine kinase protein.

KW Receptor-type tyrosine kinase; ligand; coomassie staining; PAS staining;  
 KM human.

OS Homo sapiens.

PN JP08188596-A.

PD 23-JUL-1996.

PF 13-JAN-1995; 95JP-0003677.

PR 03-NOV-1994; 94JP-0275411.

PR 19-OCT-1994; 94JP-0253848.

PA (ASAH ) ASAH KASEI KOGYO KK.

DR WPI; 1996-388601/39.

XX New ligand for receptor type tyrosine kinase - has mol.wt. 22-25  
 PT kilodalton(s) and is positive for Coomassie and PAS staining

PS Claim 5; Page 43; 51pp; Japanese.

XX AAM06333 and AAM06334 represent receptor-type tyrosine kinase receptor  
 CC binding ligands of the invention. These sequences both have the  
 CC N-terminal sequence represented by AAM06332. The ligands of the  
 CC invention recognise the fragment of the receptor type kinase receptor  
 CC represented by AAM06330 (see AAM06331 for full length sequence). The  
 CC proteins of the invention have a molecular weight of 23500 (plus or minus  
 CC 1500) Da, and are positive for Coomassie staining and PAS staining. The  
 CC protein is a new ligand of receptor-type tyrosine kinases, and can be  
 CC prepared by standard recombinant techniques.

XX Sequence 195 AA;

Query Match 23.3%; Score 431.5; DB 17; Length 195;

Best Local Similarity 45.3%; Pred. No. 1.7e-27;

Matches 87; Conservative 32; Mismatches 66; Indels 7; Gaps 4;

QY 30 LEPYVNSANKRFQAEAGVYLPQIGDRLLDLCPRARPSPHSSPNVEFYKLYLVGAAG 89

DB 5 LEPIYVNSNKKFLPGQGLVLPQIGDKLIIICPKV--DSKTVGQYEVYVWVDKQ 61

QY 90 RRCAPNPAPNLLCDBPDLDLRTIKFOEYSPYLKMGHERSHDYIITSDGTREGLE 149

DB 62 DRCITKENTPLNCARPDQDIKTIKFOEYSPYLKMGHERSHDYIITSDGTREGLE 121

QY 150 SLOGGVCLTRGMKVLLRVGQ--SPRGAAVPRKPVSEMPER-DRGAHSLDPGKENVLPD 206

DB 122 NQGGVCLTRGMKVLLRVGQDASAGSTRNKDPTRPELEAGTNGRSTSPFVKPNPGS 181

QY 207 PTSNATSRGAG 218  
 DB 182 ST-DGNSAGHSG 192

## RESULT 29

AAW11307  
 ID AAW11307 standard; protein; 195 AA.

XX AAW11307;

XX 21-MAR-1997 (first entry)

XX Receptor-type tyrosine kinase #3.

XX Receptor-type tyrosine kinase; LPM medium; ligand; autophosphorylation;  
 KW insect cell; animal cell; growth promoter; undifferentiated cell;  
 KW insulin; interleukin; fibroblast growth factor; hepatocyte growth factor;  
 KW nerve growth factor; interferon-gamma; tumour necrosis factor; inducer.

OS Homo sapiens.

PN JP08280385-A.

PD 29-OCT-1996.

PR 13-APR-1995; 95JP-0087878.

PR 13-APR-1995; 95JP-0087878.

PA (ASAH ) ASAH KASEI KOGYO KK.

DR WPI; 1997-014848/02.

PT LPM cell culture medium - for culturing cells which produce  
 PT receptor-type tyrosine kinase ligand

PS Disclosure; Page 38-39; 45pp; Japanese.

XX AAW11307 and AAW11308 represent receptor-type tyrosine kinases. These  
 CC sequences can be included in the LPM medium of the invention. The  
 CC medium is used to culture cells capable of producing a ligand for  
 CC receptor-type tyrosine kinase. The ligand serves as a physiologically  
 CC active substance to induce enzyme activity and autophosphorylation of  
 CC tyrosine kinase, which is involved in the development and differentiation  
 CC of animal and insect cells. The ligand can also be used as in a  
 CC pharmaceutical preparation for promoting growth of undifferentiated  
 CC cells. The medium may also include an inducer, such as insulin, an  
 CC interleukin, a fibroblast growth factor, or interferon-gamma.

SO Sequence 195 AA;

Query Match 23.3%; Score 431.5; DB 18; Length 195;

Best Local Similarity 45.3%; Pred. No. 1.7e-27;  
 Matches 87; Conservative 32; Mismatches 66; Indels 7; Gaps 4;

QY 30 LEPTVMSANKRFGAEGYVLPQIGRLDLCPRARPPGPHSSPNVEFYKLYLVGAAG 89

DB 5 LEPTVMSNSKFLPGGLVLPQIGRLDLCPRARPPGPHSSPNVEFYKLYLVGAAG 61

QY 90 RRCAPAPAPNLLTCRPPDLRRTIKFOEYSPNLMGHEFRSHDYIITSGTEGLE 149

DB 62 DRCTIKKENPFLNCAKPPDDIKTIKQERSPNLMGHEFRSHDYIITSGTEGLE 121

QY 150 SLGGVCLTRGMKVLKVGQ--SPRGAIVPKVSEMPMER-DRGAASLEPGKENIPGD 206

DB 122 NQEGVCCOTRMLKIMKVGDPASAGSTRNKDPTRRPELAGTNGRSSTSPVKNPGS 181

QY 207 PTSNATSRGAG 218

DB 182 ST-DGNSAGHSG 192

## RESULT 30

AAW37671  
 ID AAW37671 standard; Protein; 92 AA.

XX AAW37671;

XX 17-OCT-2001 (first entry)

XX Peptide #11708 encoded by probe for measuring placental gene expression.

XX Probe; microarray; human; placenta; antenatal diagnosis;  
 KW genetic disorder.

XX Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta -

XX Claim 27; SEQ ID No 37940; 654pp; English.

PS The present invention relates to single exon nucleic acid probes (SENP;

XX see A131315-A157546). The present sequence is a peptide encoded by one

XX such probe. The probes are useful for producing a microarray for

XX predicting, measuring and displaying gene expression in samples derived

XX from human placenta. The probes are useful for antenatal diagnosis of

XX human genetic disorders.

SO Sequence 92 AA;

Query Match 15.4%; Score 284.5; DB 22; Length 92;

Best Local Similarity 54.2%; Pred. No. 7.4e-16;  
 Matches 52; Conservative 15; Mismatches 24; Indels 5; Gaps 1;

QY 42 FOAEGYVLPQIGRLDLCPRARPPGPHSSPNVEFYKLYLVGAAGRCAPAPNLL 101

DB 1 FLSGKGLVLPKIGRLDLCPRARPPGPHSSPNVEFYKLYLVGAAGRCAPAPNLL 55

QY 102 LTCRPPDLRRTIKFOEYSPNLMGHEFRSHDYI 137

DB 56 VTCNRPQEQIRFTIKFOEYSPNLMGHEFRSHDYI 91

## RESULT 31

ABG46524  
 ID ABG46524 standard; Peptide; 92 AA.

XX ABG46524;

XX 19-AUG-2002 (first entry)

XX Human peptide encoded by genome-derived single exon probe SEQ ID 36189.

XX Human; single exon probe; asthma; lung cancer; COPD; ILD;

KM chronic obstructive pulmonary disease; interstitial lung disease;  
 KM familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KM tubercous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KM Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KM pulmonary histiocytosis; lymphangioleiomyomatosis; Karagenen syndrome;  
 KM pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KM primary ciliary dyskinesia; pulmonary hypertension;  
 KM hyaline membrane disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200186003-A2.  
 XX  
 PD 15-NOV-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00665.  
 XX  
 PR 04-FEB-2000; 2000US-180312P.  
 XX  
 PR 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2002-114183/15.  
 XX  
 PT Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples -  
 XX  
 PS Claim 27; SEQ ID No 36189; 634bp; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridize at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene  
 CC expression analysis, and for identifying exons in a gene particularly  
 CC using human lung derived mRNA and for the study of lung diseases  
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 CC fibrosis, neurofibromatosis, tubercous sclerosis, Gaucher's disease,  
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
 CC pulmonary alveolar proteinosis, Karagenen syndrome, fibrocystic  
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
 CC and hyaline membrane disease. The present sequence is a peptide/protein  
 CC encoded by a single exon probe of the invention.  
 CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 92 AA;  
 SQ  
 Query Match 15.4%; Score 284.5; DB 23; Length 92;  
 Best Local Similarity 54.2%; Pred. No. 7.4e-16;  
 Matches 52; Conservative 15; Mismatches 24; Indels 5; Gaps 1;  
 QY 42 FQAGGYVTPQIGDRULDCPPAPPPHSSPNFFYLYLVGAGGRCAPAPNLL 101  
 DB 1 FLGKGLVTPKIGDKLDLDCPPAEPGRP-----YEVYLYLVPRQAAAGCTVLDPNVL 55  
 QY 102 LDCRPPDLDFRTIKFOEYSPNLMGHEFRSHDPYI 137  
 DB 56 VTCNRPQEQIRFTIKFOEYSPNLMGHEFRSHDPYI 91  
 RESULT 32  
 ID AAM37534 standard; Protein; 136 AA.  
 AC AAM37534;  
 DT 17-OCT-2001 (first entry)  
 DE Peptide #11571 encoded by probe for measuring placental gene expression.  
 KM Probe; microarray; human; placenta; antenatal diagnosis;  
 KM genetic disorder.  
 OS Homo sapiens.  
 XX  
 PN WO200157272-A2.  
 PD 09-AUG-2001.  
 PF 30-JAN-2001; 2001WO-US00663.  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488897/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human placenta -  
 XX  
 PS Claim 27; SEQ ID No 37803; 654bp; English.  
 XX  
 CC The present invention relates to single exon nucleic acid probes (SENP:  
 CC see A113315-A1157546). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for producing a microarray for  
 CC predicting, measuring and displaying gene expression in samples derived  
 CC from human placenta. The probes are useful for antenatal diagnosis of  
 CC human genetic disorders.  
 XX  
 SQ Sequence 136 AA;  
 Query Match 11.1%; Score 205.5; DB 22; Length 136;  
 Best Local Similarity 37.7%; Pred. No. 3.5e-09;  
 Matches 52; Conservative 13; Mismatches 34; Indels 39; Gaps 5;  
 QY 230 GAAGGLA-----LILLGVAGAGA-----MCMRRRRKPSBSHHPG 266



```

Db 11 GASGSSGDDPDGFRNSKALFAAVGAGCVIFLLIIFITVLLKLRGRKHRTQO----- 65
Qy 267 GSFGRGSLGT-----GGGGGNGPREAPBEGELIALRGGAADPFCHYKVSQDYGHPV 322
Db 66 ----RAAALSLSTLASPRKSGGTAGTEPSDIIPLR---TTENNYCPHYKVSQDYGHPV 118
Qy 323 YIVDGPPOSPPNIYKYV 340
Db 119 YIVDGPPOSPPNIYKYV 136

RESULT 33
ABG46394
ID ABG46394 standard; Peptide; 136 AA.
XX
XX ABG46394;
XX
XX 19-AUG-2002 (first entry)
XX
XX Human peptide encoded by genome-derived single exon probe SEQ ID 36059.
XX
XX Human; single exon probe; asthma; lung cancer; COPD, IHD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease.
XX
XX Homo sapiens.
XX
XX MO200186003-A2.
XX
XX 15-NOV-2001.
XX
XX 30-JAN-2001; 2001WC-US00665.
XX
XX 04-FEB-2000; 2000US-180312P.
XX 26-MAY-2000; 2000US-207456P.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-234687P.
XX 27-SEP-2000; 2000US-236359P.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2002-114183/15.
XX
XX Spatially-addressable set of single exon nucleic acid probes, used to
XX measure gene expression in human lung samples -
XX
XX Claim 27; SEQ ID No 36059; 634bp; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human lung comprising single exon nucleic acid probes having one of
XX 12614 nucleic acid sequences mentioned in the specification, or their
XX complements or the 12387 open reading frames derived from the 12614
XX probes. Also included are a microarray comprising the novel set of
XX probes; the novel set of probes which hybridise at high stringency to a
XX nucleic acid expressed in the human lung; measuring gene expression in a
XX sample derived from human lung, comprising (a) contacting the array with
XX mRNA, and (b) measuring the label detectably bound to each probe of
XX the array; identifying exons in a eukaryotic genome, comprising
XX (a) algorithmically predicting at least one exon from genomic sequences
XX of the eukaryote; and (b) detecting specific hybridisation of detectably

```

```

CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene
CC expression analysis, and for identifying exons in a gene, particularly
CC using human lung derived mRNA and for the study of lung diseases
CC such as asthma, lung cancer, chronic obstructive pulmonary disease
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
CC pulmonary alveolar proteinosis, Karsenger syndrome, fibrocystic
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
CC and hyaline membrane disease. The present sequence is a peptide/protein
CC encoded by a single exon probe of the invention.
CC
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pcl_sequences.
XX
XX SQ Sequence 136 AA;
XX
XX Query Match 11.1%; Score 205.5; DB 23; Length 136;
XX Best Local Similarity 37.7%; Pred. No. 3.5e-09;
XX Matches 52; Conservative 13; Mismatches 34; Indels 39; Gaps 5;
XX
Qy 230 GAAGGA-----LLLVAGAGA-----MCRRRRAKPSRRHGP 266
Db 11 GASGSSGDDPDGFRNSKALFAAVGAGCVIFLLIIFITVLLKLRGRKHRTQO----- 65
Qy 267 GSFGRGSLGT-----GGGGGNGPREAPBEGELIALRGGAADPFCHYKVSQDYGHPV 322
Db 66 ----RAAALSLSTLASPRKSGGTAGTEPSDIIPLR---TTENNYCPHYKVSQDYGHPV 118
Qy 323 YIVDGPPOSPPNIYKYV 340
Db 119 YIVDGPPOSPPNIYKYV 136

RESULT 34
AAV71437
ID AAV71437 standard; peptide; 82 AA.
XX
XX AAV71437;
XX
XX 04-OCT-2000 (first entry)
XX
XX Human ephrin B2 C-terminal cytoplasmic domain.
XX
XX PDZ domain; B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;
XX cellular process; nerve cell interaction; regeneration of nerve cell;
XX axonogenesis; antiproliferative; proliferative disorder; treatment;
XX differentiative disorder; human; cytoplasmic domain.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH Region 50..68
XX FT /note= "Conserved region"
XX FT Modified-site 53
XX FT /note= "Conserved tyrosine phosphorylation site"
XX FT Modified-site 60
XX FT /note= "Conserved tyrosine phosphorylation site"
XX FT Modified-site 65
XX FT /note= "Conserved tyrosine phosphorylation site"
XX FT Region 71..75

```

```

FT      /note= "Conserved region"
FT      Region 77..82
FT      /note= "Conserved region"
FT      Domain 79..82
FT      /label= PDZ_domain
FT      Modified-site 79
FT      /note= "Conserved tyrosine phosphorylation site"
FT      Modified-site 80
FT      /note= "Conserved tyrosine phosphorylation site"
FT      /note= "Conserved tyrosine phosphorylation site"
XX      WO200031124-A2.
XX      PD 02-JUN-2000.
XX      PF 19-NOV-1999; 99WO-CA01101.
XX      PR 20-NOV-1998; 98US-0109158.
XX      PA (MOUN ) MOUNT SINAI HOSPITAL.
XX      PI Lin D, Pawson A;
XX      DR WPI, 2000-400038/34.
XX      PT Isolated complex for treating proliferative or differentiative
XX      disorders comprises B class ephrin and PDZ domain containing protein -
XX      Example; Fig 1; 59pp; English.
XX      CC The patent discloses a complex comprising of a B class ephrin and PDZ
XX      CC domain containing protein. B class ephrins function as ligands for Eph
XX      CC receptor tyrosine kinases (RTK) and possess a transmembrane element and a
XX      CC highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ
XX      CC binding site. This complex is used in methods to modulate the interaction
XX      CC of a B class ephrin and PDZ domain containing protein and to identify
XX      CC modulator compounds. It is also used for modulating cellular processes
XX      CC like, axonogenesis, nerve cell interactions and regeneration of nerve
XX      CC cells. The complex is also useful for treating proliferative or
XX      CC differentiative disorders associated with this protein complex.
XX      CC The present sequence is the human ephrin B2 C-terminal cytoplasmic
XX      CC domain, comprising conserved sequences reminiscent of known or predicted
XX      CC binding sites for PDZ domains. Ephrin B2 is also known as Hck-L, ERF-2,
XX      CC LERK-5 and NLERK-1.
XX      CC LERK-5 and NLERK-1.
XX      SQ Sequence 82 AA;
XX      QY Query Match 11.0%; Score 204; DB 21; Length 82;
XX      DB Best Local Similarity 66.1%; Pred. No. 2.5e-09;
XX      DB Matches 41; Conservative 4; Mismatches 11; Indels 6; Gaps 2;
XX      QY 279 GGGGMPREAPGELGIALRGGAADPPCPHYEKVSGDYGHPVYIVDGPSPNNIYK 338
XX      DB 27 GGNNG--SEPSDVIPLR---TADSVCPHYEKVSGDYGHPVYIVQMPPOSPANNIY 80
XX      QY 339 KV 340
XX      DB 81 KV 82
XX      AC AAB54187;
XX      AC AAB54187;
XX      DT 09-MAR-2001 (first entry)
XX      DE Human pancreatic cancer antigen protein sequence SEQ ID NO:639.
XX      KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
XX      KW detection; diagnosis; identification; cytoskeletal; neuroprotective;
XX      KW neurotropic; immunomodulatory; relaxant; contractile; gynaecological;
XX      KW antiinflammatory; cardiac; gene therapy; chromosome mapping;

```

```

XX      KW linkage analysis; tissue identification; tissue typing; forensic;
XX      KW neural; immune system; muscular; reproductive; gastrointestinal;
XX      KW pulmonary; cardiovascular; renal; proliferative.
XX      OS Homo sapiens.
XX      PN WO200055320-A1.
XX      PD 21-SEP-2000.
XX      PF 08-MAR-2000; 2000WO-US05589.
XX      PR 12-MAR-1999; 99US-0124270.
XX      PA (HUMA-) HUMAN GENOME SCI INC.
XX      PI Rosen CA, Ruben SM;
XX      DR WPI; 2000-579444/54.
XX      DR N-PSDB; AAC98952.
XX      PT New nucleic acid that is a pancreatic cancer antigen for preventing,
XX      PT treating, or ameliorating a medical condition, particular pancreatic
XX      PT cancer, or for use in assays for diagnosing a pathological condition -
XX      Claim 11; Page 1078; 1379pp; English.
XX      CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
XX      CC proteins, called pancreatic cancer antigens, given in AAB54008 to
XX      CC AAB54466. The human pancreatic cancer antigens have cytoskeletal,
XX      CC neuroprotective, neurotropic, immunomodulatory, relaxant, contractive,
XX      CC gynaecological, cardiac and antiinflammatory activities, and can be used
XX      CC in gene therapy. The polynucleotide and proteins can be used for
XX      CC preventing, treating, or ameliorating a medical condition or in assays
XX      CC for diagnosing a pathological condition and a susceptibility to one in a
XX      CC subject. Binding partners to the proteins and the activity of the
XX      CC proteins can be identified. The pancreatic cancer antigens can be used to
XX      CC detect, treat or prevent pancreatic disorders, especially cancer.
XX      CC Apoptosis and antagonists to the antigens can be screened for. The
XX      CC pancreatic cancer antigen polynucleotides can be used to design nucleic
XX      CC acid hybridisation probes that can be used in chromosome mapping, linkage
XX      CC analysis, tissue identification and/or typing and a variety of forensic
XX      CC and diagnostic methods. The proteins can be used to generate antibodies
XX      CC which are used to purify, detect and target the polypeptides, including
XX      CC both in vivo and in vitro diagnostic and therapeutic methods. The
XX      CC proteins can be used to treat or prevent neural, immune system, muscular,
XX      CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
XX      CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
XX      CC sequences used in the exemplification of the present invention.
XX      SQ Sequence 106 AA;
XX      QY Query Match 10.9%; Score 202.5; DB 21; Length 106;
XX      DB Best Local Similarity 43.6%; Pred. No. 4.5e-09;
XX      DB Matches 48; Conservative 11; Mismatches 26; Indels 25; Gaps 4;
XX      QY 235 LALLLVAGAGAGCMCRRRRAKPSBSRHPGSGSLGJ---GGGGGMPRAEP 220
XX      DB 18 LTVLLRLK-----RKRRKRTQO-----RAALSLSTLSPRGSGGTACTEP 59
XX      QY 291 GELGIALRGGAADPPCPHYEKVSGDYGHPVYIVDGPSPNNIYK 340
XX      DB 60 SDIILPLR---TENNCPHYEKVSGDYGHPVYIVQMPPOSPANNIYK 106
XX      AC AAY71436;
XX      AC AAY71436;
XX      DT 04-OCT-2000 (first entry)
XX      DE Human pancreatic cancer antigen protein sequence SEQ ID NO:639.
XX      KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
XX      KW detection; diagnosis; identification; cytoskeletal; neuroprotective;
XX      KW neurotropic; immunomodulatory; relaxant; contractile; gynaecological;
XX      KW antiinflammatory; cardiac; gene therapy; chromosome mapping;

```

DE Human ephrin B1 C-terminal cytoplasmic domain.  
 XX PDZ domain; B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;  
 KW cellular process; nerve cell interaction; regeneration of nerve cell;  
 KW axonogenesis; antiproliferative; proliferative disorder; treatment;  
 KW differentiative disorder; human; cytoplasmic domain.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 50..68  
 FT /note= "Conserved region"  
 FT Modified-site 53  
 FT /note= "Conserved tyrosine phosphorylation site"  
 FT Modified-site 60  
 FT /note= "Conserved tyrosine phosphorylation site"  
 FT Modified-site 65  
 FT /note= "Conserved tyrosine phosphorylation site"  
 FT Region 71..75  
 FT /note= "Conserved region"  
 FT Region 77..82  
 FT /note= "Conserved region"  
 FT Domain 79..82  
 FT /label= PDZ domain  
 FT Modified-site 79  
 FT /note= "Conserved tyrosine phosphorylation site"  
 FT Modified-site 80  
 FT /note= "Conserved tyrosine phosphorylation site"  
 FT Modified-site /note= "Conserved tyrosine phosphorylation site"  
 XX  
 XX W0200031124-A2.  
 XX  
 XX 02-JUN-2000.  
 XX  
 XX 19-NOV-1999; 99WO-CA01101.  
 XX  
 XX 20-NOV-1998; 98US-0109158.  
 XX  
 XX (MOUN ) MOUNT SINAI HOSPITAL.  
 XX  
 XX Lin D, Pawson A;  
 XX  
 XX WPI; 2000-400038/34.  
 XX  
 XX Isolated complex for treating proliferative or differentiative  
 PT disorders comprises B class ephrin and PDZ domain containing protein -  
 XX  
 XX Example; Fig 1; 59pp; English.  
 XX  
 CC The patent discloses a complex comprising of a B class ephrin and PDZ  
 CC domain containing protein. B class ephrins function as ligands for Eph  
 CC receptor tyrosine kinases (RTK) and possess a transmembrane element and a  
 CC highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ  
 CC binding site. This complex is used in methods to modulate the interaction  
 CC of a B class ephrin and PDZ domain containing protein and to identify  
 CC modulator compounds. It is also used for modulating cellular processes  
 CC like, axonogenesis, nerve cell interactions and regeneration of nerve  
 CC cells. The complex is also useful for treating proliferative or  
 CC differentiative disorders associated with this protein complex.  
 CC The present sequence is the human ephrin B1 C-terminal cytoplasmic  
 CC domain, comprising conserved sequences reminiscent of known or predicted  
 CC binding sites for PDZ domains. Ephrin B1 is also known as LERK-2, Elk-L,  
 CC BFL-3, Cck-L and STRA1.  
 CC  
 XX  
 SQ Sequence 82 AA;  
 Query Match 10.9%; Score 201; DB 21; Length 82;  
 Best Local Similarity 47.3%; Pred. No. 4,4e-09;  
 Matches 44; Conservative 9; Mismatches 24; Indels 16; Gaps 3;  
 QY 252 RRRRAKPSBSHPPGSGRGSGGLT-----GGGGMPPRRRPEBLLDIALRGGAADPPF 307  
 DB 2 RRRRRKHTQQ-----RAAALSLTSLASPKGSGTAGTEPDDITIPR---TTENNY 49

QY 308 CPYKESGDYGHVPYIVODGPGSPNITYKV 340  
 DB 50 CPYKESGDYGHVPYIVQENPQSPNITYKV 82  
 RESULT 37  
 ID ABB58621  
 XX ABB58621 standard; Protein; 652 AA.  
 XX  
 AC ABB58621;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 2655.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN W0200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 XX 23-MAR-2001; 2001WO-US09231.  
 XX  
 XX 23-MAR-2000; 2000US-191637P.  
 XX  
 XX 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 XX WPI; 2001-656860/75.  
 XX  
 XX N-PSDB; ABL02724.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Disclosure; SEQ ID NO 2655; 21pp + Sequence listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
 CC sequences (AB01840-AB16175) and the encoded proteins  
 CC (AB57737-AB572072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX  
 SQ Sequence 652 AA;  
 Query Match 9.8%; Score 182; DB 22; Length 652;  
 Best Local Similarity 22.9%; Pred. No. 1.9e-06;  
 Matches 88; Conservative 50; Mismatches 136; Indels 110; Gaps 18;  
 QY 33 VYVNSANKRFOAG-GVLYVPOIG-----DRLDLCPRAPPGHSPNTEFYLYLVG 85  
 DB 219 VYVNTSNTSIFRIDNTDHIIDVAKGTLAFEPQVHIC-VYEPGFHEET-EKYLIVVS 276  
 QY 86 GAQGRCEAPPA-PULLLTCDPDDIDLRFTTKFQESYSLNMGHFRSHDDYIIITSQGT 144  
 DB 277 KVEYTCRITVNDPVIACDPQCLMEFTITFRPFPQGLBFLPNDYFIFISTS--S 334  
 QY 145 RRGLESLOGGVLTGRGMVILRVGSPRGAVPRPVSMEWERRGAHSL-----EPG 199  
 DB 333 KODLYRRIGRCSTNNKAVFRKVCAPEDNN-KTTALNSKSVYDTGAINVANNANDES 393  
 QY 200 KENLPDPTSNATSRGAG-----PLPP-----P 223

```

Db      394 HVNHSNNIAIGTNGGQIIGPQSAQIPNPLSGNNNINGIPTTNSNDQFNRI 453
Qy      224 SMPAVGAGAGLALLLVAGAGAMCMRRRAKPSRHPGP-SFGGSLGIGGGG 282
Db      454 IQPNIIGNHVGNVAGTGVGGGIIIL-----TPGAHGNINMLQPRGGI 499
Qy      283 MGRREAPG---ELGIALRG-----GGAADPPFCPHYEKVSGDYG 320
Db      500 NG---AYPGHHIQTGIRINNVPTQHNYP SHKGNANSNINGNDH---HHYK-----H 547
Qy      321 PYTIYVD-----GPPQSPENIY 337
Db      548 PNEVKNBELTNSGATSDGNIF 571

```

RESULT 38  
ABR65548  
ID ABR65548 standard; Protein; 652 AA.

AC ABR65548;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 23436.

KM Drosophila: developmental biology; cell signalling; insecticide;  
pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

PT N-PSDB; ABL09661.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
interactions.

PS Disclosure; SEQ ID NO 23436; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABR57737-ABR72072).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 652 AA;

XX

Query Match 9.8%; Score 182; DB 22; Length 652;

Best Local Similarity 22.9%; Pred No. 1.9e-06;

Matches 88; Conservative 50; Mismatches 136; Indels 110; Gaps 18;

Qy 33 VYNSANKRQAGG-GVLYPQIG-----DRIDLCPARPDPGPHSSPNVYFKLYLVG 85

```

Db      219 MHNTNSISIRINTDHIIDVKNLAFFEDVHICP-VYEGTRENET-EKTIYVNS 276
Qy      86 GAQRRCEAPPA-PNLLTCRDPDLRAFTIKQEXSPNLMGHEFRSHDDYIIATSDGT 144
Db      277 XVEYTCRIINADPRVIALCDKPOKLMFTITFRPTPOPGGLEFPGNDYFISTSS 334
Qy      145 REGLESIGGVCTLRGMKYLAVGQSPRGGAVERKVSSEMPERDGAHSL-----EPG 199
Db      335 KDLRYRIGGRCSTNNKRVKVCAPEDNN-KTYALNSKSVYTDGALINVIANNDES 393
Qy      200 KENLPQDPTSNATSRGAEG-----PLFP-----P 223
Db      394 HVNHSNNIAIGTNGGQIIGPQSAQIPNPLSGNNNINGIPTTNSNDQFNRI 453
Qy      224 SMPAVGAGAGLALLLVAGAGAMCMRRRAKPSRHPGP-SFGGSLGIGGGG 282
Db      454 IQPNIIGNHVGNVAGTGVGGGIIIL-----TPGAHGNINMLQPRGGI 499
Qy      283 MGRREAPG---ELGIALRG-----GGAADPPFCPHYEKVSGDYG 320
Db      500 NG---AYPGHHIQTGIRINNVPTQHNYP SHKGNANSNINGNDH---HHYK-----H 547
Qy      321 PYTIYVD-----GPPQSPENIY 337
Db      548 PNEVKNBELTNSGATSDGNIF 571

```

RESULT 39

ID AAR82605 standard; Protein; 234 AA.

AC AAR82605;

DT 16-MAY-1996 (first entry)

DE Eph transmembrane tyrosine kinase family ligand, Efl-2.

KM Efl-2; EHK1-L; Eph transmembrane tyrosine kinase family ligand;

XX neurological disorder; identification; diagnosis.

OS Homo sapiens.

PI Key Location/Qualifiers

FT Peptide 1..30

FT Region 218..235 /label= signal\_peptide

FT /note= "carboxy terminal hydrophobic

FT Misc-difference 158 /note= "residue borders main conserved regions"

PN WO9527060-A2.

PD 12-OCT-1995.

PF 04-APR-1995; 95WO-US04208.

PR 21-OCT-1994; 94US-0327423.

PR 04-APR-1994; 94US-0222075.

PR 12-APR-1994; 94US-0229402.

PR 01-SEP-1994; 94US-0299567.

XX (REGF-) REGENERON PHARM INC.

XX Aldrich TH, Davis S, Gale N, Goldfarb M, Maisompierre PC;

XX Yancopoulos GD;

XX WPI; 1995-358635/46.

DR N-PSDB; AAT03883.

XX ligands which bind Eph family receptors - used in the diagnosis of  
PT neurological disorders

PS Disclosure; Fig 2; 58pp; English.

XX Efl-2 (also known as EHK-1L) is an Eph transmembrane tyrosine kinase

CC family ligand. It has homology with Etk (Etk-1) (see AA02604). Efl-2

CC ends in a C-terminal hydrophobic sequence that appears to be a

CC recognition sequence allowing it to be GPI-linked and thus lacking in

CC an intracellular domain. Efl-2 is useful for identifying other ligands

CC for Etk-1, -2, -3, Etk and Etk receptors. The ligands are useful in

CC promoting a differential function and/or influencing the phenotype,

CC such as growth and/or proliferation, of receptor bearing cells. They

CC may be used in the diagnosis, and treatment of neurological disorders.

XX

SQ Sequence 234 AA;

Query Match 9.8%; Score 181; DB 16; Length 234;

Best Local Similarity 28.4%; Pred. No. 6.7e-07;

Matches 64; Conservative 23; Mismatches 82; Indels 56; Gaps 11;

QY 7 GPGGVVAVALLLVGLVGLSLBPYVNSANKRFOAEGYLYPGIGRDLDCP--R 64

DB 24 GPG-----GALG----NRHAVYVNSNQHLRRE-GYVQVNVNDYLDIYCFHYN 67

QY 65 ARPPGPHSSPNVEFYKLYLVGAQGRCEAPAPNILLTCDRPDL--DLRPTIKFQEYS 121

DB 68 SSGAGPDPGGAGQYLYVNSRNGYRTCNASQFK-RMECNRPAAHSPIKSEKQRTS 126

QY 122 PNIWGEFRSHDHYIATSDGTREGLESIGGVCLTRGKVLRLVQSPRGAVPRKPV 181

DB 127 AFSIGYFFHAGHEYYIISTPTNHL-----WKCLR--MKVFCVCASTSHSG---EKV 174

QY 182 SEMP-----MERDGAHSLF-----PKENTLP 204

DB 175 PTFPTMGPNVKINLEDFEGENPQVPLEKSIQSTSPRREHL 219

RESULT 40

AA071481

ID AA071481 standard; Protein; 238 AA.

XX

AC AA071481;

XX

DT 03-OCT-1995 (first entry)

XX

DE Human hek-L protein.

XX

KW ligand; cell surface; tyrosine kinase receptor; tumorigenesis;

KW immunogen.

XX

OS Homo sapiens.

XX

FM Key Location/Qualifiers

FT Peptide 1..19

FT Protein /note="signal peptide" 20..238

XX

PN WO9506065-A.

XX

PD 02-MAR-1995.

XX

PF 17-AUG-1994; 94WO-US09282.

XX

XX 20-AUG-1993; 93US-0109745.

XX 30-AUG-1993; 93US-0114426.

XX 03-DEC-1993; 93US-0161132.

XX 09-MAY-1994; 94US-0240124.

XX

PA (IMMV ) IMMUNEX CORP.

XX

PI Beckmann MP, Cerretti DP;

XX

DR WPI; 1995-106811/14.

XX

DR N-PSDB; AA085887.

XX

PT New isolated DNA encoding hek-L protein or its fusion products -

PT useful as assay reagent or for carrying therapeutic and

PT diagnostic compounds to leukemia cells.

XX

PS Claim 21; Page 36; 45pp; English.

XX

CC The sequence is that of a novel protein designated hek-L, a protein

CC that can bind hek (a cell surface receptor tyrosine kinase). Hek-L

CC is the first known ligand for hek and can be used to study cellular

CC processes regulated by hek (which may be involved in tumorigenesis).

CC It is also an immunogen for antibody production, as a reagent for

CC detecting hek or hek-L in in vitro assays, to determine binding of

CC hek proteins, to purify hek proteins, and to carry diagnostic or

CC cytotoxic agents to particular leukemia cells that express the hek

CC antigen. Hek-L also binds the elk tyrosine kinase receptors.

CC See also AA071482.

XX

SQ Sequence 238 AA;

Query Match 9.7%; Score 179; DB 16; Length 238;

Best Local Similarity 28.4%; Pred. No. 1e-06;

Matches 65; Conservative 24; Mismatches 80; Indels 60; Gaps 12;

QY 7 GPGGVVAVALLLVGLVGLSLBPYVNSANKRFOAEGYLYPGIGRDLDCP--R 64

DB 24 GPG-----GALG----NRHAVYVNSNQHLRRE-GYVQVNVNDYLDIYCFHYN 67

QY 65 ARPPGPHSSP---NVEFYKLYLVGAQGRCEAPAPNILLTCDRPDL--DLRPTIKF 117

DB 68 SSGVFGAGPDPGGAGQYLYVNSRNGYRTCNASQFK-RMECNRPAAHSPIKSEK 126

QY 118 QEYSPNIMGEFRSHDHYIATSDGTREGLESIGGVCLTRGKVLRLVQSPRGAVP 177

DB 127 QRYAFSLGYEFFHAGHEYYIISTPTNHL-----WKCLR--MKVFCVCASTSHSG--- 174

QY 178 RKPVSEMP-----MERDGAHSLF-----PKENTLP 204

DB 175 EKVPTLPQFTMGPNVKINLEDFEGENPQVPLEKSIQSTSPRREHL 223

RESULT 41

AA02587

ID AA02587 standard; Protein; 184 AA.

XX

AC AA02587;

XX

DT 28-NOV-1996 (first entry)

XX

DE Lerk-6 protein.

XX

KW Lerk-6; hek; elk; cell surface receptor; culture; reagent;

KW neuron; disorder; injury; delivery agent; diagnostic; therapeutic;

KW Lerk-7; probe; cytokine.

XX

OS Mus musculus.

XX

PN WO9617925-A1.

XX

PD 13-JUN-1996.

XX

PF 05-DEC-1995; 95WO-US15781.

XX

XX 01-MAR-1995; 95US-0396946.

XX 06-DEC-1994; 94US-0351025.

XX

PA (IMMV ) IMMUNEX CORP.

XX

PI Cerretti DP;

XX

DR WPI; 1996-287171/29.

XX

DR N-PSDB; AA032700.

XX

PT New isolated human Lerk-7 cytokine - which binds to cell surface

PT and cell surface screening procedures

production of the proteins. LERK-6 polypeptides may be useful in the









QY 123 NLWGHEFRSHDYIATLS 141  
 DB 123 FSLGFEFRPGHYYIATLS 141

## RESULT 49

AAW71007  
 ID AAW71007 standard; Protein; 200 AA.

XX AC AAW71007;

XX DT 20-OCT-1998 (first entry)

XX DE Amino acid sequence of an avian E1f-1 protein.

XX KW Chicken; EPH receptor ligand; E1f-1; mek-4; sek-AF;  
 KW tyrosine kinase ligand; B61; LERK-2; proliferation; differentiation;  
 KW intracellular signalling; increased; survival; neuronal cell;  
 KW neuron survival; treatment; Alzheimer's; Parkinson's; lymphatic tumour;  
 KW artificial liver; cartilage; bone formation.

XX OS Gallus sp.

XX FH Key Location/Qualifiers

XX FT Misc-difference 61..150

XX FT Note="contains a Cys4 motif"

XX FT Misc-difference 35..157

XX FT Note="contains a core sequence motif"

XX PN US5795734-A.

XX PD 18-AUG-1998.

XX PF 31-MAY-1995; 95US-0455001.

XX PR 31-MAY-1995; 95US-0455001.

XX PR 19-SEP-1994; 94US-0308814.

XX PR 27-FEB-1995; 95US-0393462.

XX PA (HARD) HARVARD COLLEGE.

XX PI Cheng H, Flanagan JG;

XX DR WPI; 1998-46665/40.

XX DR N-PSDB; AAV42927.

XX PT Nucleic acid encoding E1f-1 protein that binds to EPH-type receptor

XX PT - for production of E1f-1 protein, useful for regulating

XX PT proliferation, differentiation, and survival of cells

XX PS Claim 1; Columns 75-78; 53pp; English.

XX CC The present sequence represents an avian EPH receptor ligand designated  
 CC E1f-1. This ligand can bind to both mek-4 and sek-AF. E1f-1 is a  
 CC tyrosine kinase ligand, which is linked to the membrane through a  
 CC phosphatidylinositol linkage. It shares some homology to 2 other EPH  
 CC receptor ligands, B61 and LERK-2. The E1f-1 protein modulates  
 CC proliferation, differentiation and survival of EPH receptor-expressing  
 CC cells by stimulating or antagonising intracellular signalling mediated  
 CC by the EPH receptor. Typical of many potential applications are  
 CC increasing survival of neuronal cells in culture (e.g. where intended for  
 CC transplantation), also therapeutically in increase neuron survival  
 CC (e.g. treatment of Alzheimer's or Parkinson's diseases), to prevent  
 CC nervous system and lymphatic tumours, to induce differentiation of  
 CC hepatocytes to form an artificial liver, to induce cartilage and bone  
 CC formation.

XX SQ Sequence 200 AA;

Query Match 9.3%; Score 172; DB 19; Length 200;  
 Best Local Similarity 35.3%; Pred. No. 3e-06; 58; Indels 16; Gaps 5;  
 Matches 49; Conservative 16; Mismatches 58; Indels 16; Gaps 5;

QY 15 ALLLGLVGLVSGLSLEP-----VYVNSANKRFQAEQGVLYLPQIGRLDLCPRA 65  
 DB 7 AALLAIVG-VCWSDDDGKATSDRYAVYVNSNRFH-RGDTYEVAVINDYLDYCHY 64

QY 66 RPPGHSHPNVEFYLYLVGAGRCCEAPPANLLITCDPDL---DLFTIKQEQSP 122

DB 65 EEPLP--AERMERVLYLVYVNEGHASCDHRQGFGRMECNRPDPSGDLKFEKQLFTP 122

QY 123 NLWGHEFRSHDYIATLS 141

DB 123 FSLGFEFRPGHYYIATLS 141

## RESULT 50

AAE71482  
 ID AAE71482 standard; Protein; 201 AA.

XX AC AAE71482;

XX DT 03-OCT-1995 (first entry)

XX DE Human hek-L protein.

XX KW Ligand; cell surface; tyrosine kinase receptor; tumorigenesis;  
 KW immunogen.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT Peptide 1..22

XX FT Note="signal peptide"

XX FT Protein 23..201

XX PN WO9506065-A.

XX PD 02-MAR-1995.

XX PF 17-AUG-1994; 94WO-US09282.

XX PR 20-AUG-1993; 93US-0109745.

XX PR 30-AUG-1993; 93US-0114426.

XX PR 03-DEC-1993; 93US-0161132.

XX PR 09-MAY-1994; 94US-0240124.

XX PA (IMV) IMMUNEX CORP.

XX PI Beckmann MP, Cerretti DP;

XX DR WPI; 1995-106811/14.

XX DR N-PSDB; AAO85888.

XX PT New isolated DNA encoding hek-L protein or its fusion products

XX PT useful as assay reagent or for carrying therapeutic and

XX PT diagnostic compounds to leukemia cells.

XX PS Claim 21; Page 38; 45pp; English.

XX CC The sequence is that of a novel protein designated hek-L, a protein

XX CC that can bind hek (a cell surface receptor tyrosine kinase). Hek-L

XX CC is the first known ligand for hek and can be used to study cellular

XX CC processes regulated by hek (which may be involved in tumorigenesis).

XX CC It is also an immunogen for antibody production, as a reagent for

XX CC detecting hek or hek-L in vitro assays, to determine binding of

XX CC hek proteins, to purify hek proteins, and to carry diagnostic or

XX CC cytotoxic agents to particular leukemia cells that express the hek

XX CC antigen. Hek-L also binds the elk tyrosine kinase receptors.

XX CC See also AAE71481.

XX SQ Sequence 201 AA;

Query Match 9.2%; Score 169.5; DB 16; Length 201;  
 Best Local Similarity 29.9%; Pred. No. 4.9e-06;  
 Matches 66; Conservative 18; Mismatches 82; Indels 55; Gaps 10;

QY 24 LVSGLSL-EPYVWNSANKRFOABEGYVLYPOIGRDLILCPRARPPGHSSPNVEFYKLY 82  
 Db 20 LRGSLSLHVVYWNSSNPRL-LRGDAVVELGLNDYLDIVCPHYEGPPEGP--ETFALY 76  
 QY 83 LVGGAQGRCEAP-PAPNILLTCDRPDLDLRTIKFOEYSPNLMGHEFRSHDYIATS 141  
 Db 77 MVDWPGYSCQAEGRPRAYKRWVCSLFGHVQPSKIQRTFPLSGFEFLPGETIYIISVP 136  
 QY 142 DGTREGLSLQGGVCLTRGMKVLRLVGOSPRGAVPRKPVSEMPMERDRGAHSLPEGKE 201  
 Db 137 --TPE-----SSGQCL-----RLOVSVCCCKERKESAHFPV----- 164  
 QY 202 NLPGDPTSNATS--RGAEGLPPEPMPAPVAGAGLALLL 240  
 Db 165 ---GSPGESGTSGRGDPSP-----LCILLL 189

Search completed: February 11, 2003, 12:03:00  
 Job time : 32.2264 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2003, 12:01:50 / Search time 10.6918 Seconds  
(without alignments)  
935.649 Million cell updates/sec

Title: US-10-021-121-4

Perfect score: 1850  
Sequence: 1 MGPPHSGGQGVVGVALLLIG.....PYIVQDPPGSPFNYYKV 340

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 50 summaries

Database: Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/backfillset1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1846	99.8	340	4	US-09-214-631-3
2	637.5	34.5	336	1	US-08-436-044-2
3	637.5	34.5	336	2	US-08-436-054-2
4	637.5	34.5	336	5	PCT-US95-08812-2
5	632	34.2	346	1	US-08-213-403-2
6	632	34.2	346	1	US-08-458-077-2
7	632	34.2	346	1	US-08-460-741-2
8	632	34.2	346	1	US-08-747-240-2
9	632	34.2	346	1	US-08-299-567-6
10	629.5	34.0	333	1	US-08-436-044-4
11	629.5	34.0	333	2	US-08-436-054-4
12	629.5	34.0	333	4	US-08-271-948-2
13	629.5	34.0	333	5	PCT-US95-08534-2
14	629.5	34.0	333	5	PCT-US95-08812-4
15	622.5	33.6	333	4	US-09-214-631-4
16	608.5	32.9	345	4	US-09-214-631-5
17	498	26.9	89	4	US-09-214-631-13
18	201	10.9	82	4	US-09-214-631-11
19	200.5	10.8	82	4	US-09-214-631-12
20	179	9.7	234	4	US-08-299-567-5
21	179	9.7	238	1	US-08-240-124-2
22	179	9.7	238	1	US-08-453-943-2
23	179	9.7	238	2	US-09-057-121-2
24	179	9.7	238	4	US-09-358-121-4
25	176.5	9.5	135	1	US-08-289-567-7
26	176	9.5	184	1	US-09-609-324A-2
27	176	9.5	184	2	US-08-920-440B-2

28	176	9.5	184	4	US-09-173-492-2	Sequence 2, Appli
29	176	9.5	184	4	US-09-173-133-2	Sequence 2, Appli
30	176	9.5	184	4	US-09-165-533-2	Sequence 2, Appli
31	176	9.5	184	5	PCT-US95-12779-2	Sequence 2, Appli
32	176	9.5	184	5	PCT-US95-15781-2	Sequence 2, Appli
33	176	9.5	209	1	US-08-455-001-2	Sequence 2, Appli
34	176	9.5	209	4	US-08-308-814-2	Sequence 2, Appli
35	176	9.5	209	5	US-09-11869-2	Sequence 2, Appli
36	175.5	9.5	213	1	US-09-609-324A-10	Sequence 10, Appli
37	175.5	9.5	213	2	US-08-920-440B-10	Sequence 10, Appli
38	175.5	9.5	213	4	US-09-173-492-10	Sequence 10, Appli
39	175.5	9.5	213	4	US-09-173-133-10	Sequence 10, Appli
40	174.5	9.4	233	4	US-09-214-631-7	Sequence 7, Appli
41	172	9.3	200	1	US-08-455-001-4	Sequence 4, Appli
42	172	9.3	200	5	PCT-US95-11869-4	Sequence 4, Appli
43	169.5	9.2	201	1	US-08-240-124-4	Sequence 4, Appli
44	169.5	9.2	201	1	US-08-453-943-4	Sequence 4, Appli
45	169.5	9.2	201	2	US-09-057-121-4	Sequence 4, Appli
46	169.5	9.2	201	4	US-09-358-121-4	Sequence 4, Appli
47	169	9.1	201	4	US-09-214-631-8	Sequence 8, Appli
48	168.5	9.1	209	4	US-09-214-631-6	Sequence 6, Appli
49	167.5	9.1	228	1	US-08-442-248-4	Sequence 4, Appli
50	167.5	9.1	228	1	US-08-440-815-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1  
US-09-214-631-3  
Sequence 3, Application US/09214631  
Patent No. 6413730

GENERAL INFORMATION:  
APPLICANT: Hollander, Sacha  
APPLICANT: Mbamalu, Geraldine  
APPLICANT: Pawsom, Tony  
TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR  
TITLE OF INVENTION: TYROSINE KINASES  
FILE REFERENCE: 11757.23USWO  
CURRENT APPLICATION NUMBER: US/09/214,631  
CURRENT FILING DATE: 1999-03-12  
EARLIER APPLICATION NUMBER: PCT/CA97/00473  
EARLIER FILING DATE: 1997-07-04  
EARLIER APPLICATION NUMBER: 60/021,272  
EARLIER FILING DATE: 1996-07-05  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3  
LENGTH: 340  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-214-631-3

Query Match 99.8%, Score 1846, DB 4, Length 340;  
Best Local Similarity 99.7%, Pred. No. 1.1e-146;  
Matches 339; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	MGPPHSGGQGVVGVALLLIGVLSGLSPYVYNSANKRFQARGVYVYPOIGDRDL	60
DB	1	MGPPHSGGQGVVGVALLLIGVLSGLSPYVYNSANKRFQARGVYVYPOIGDRDL	60
QY	61	LCPRARPPGPHSSPYVEFYKYLIVGAQGRCEAPAPNLLLTCDRPLDRLFTTKFOEY	120
DB	61	LCPRARPPGPHSSPYVEFYKYLIVGAQGRCEAPAPNLLLTCDRPLDRLFTTKFOEY	120
QY	121	SENWGHFRSHHDYIYIATSDGTREGLESTQGGVCLTRKVKVILRVQSPRGCAVPPKP	180
DB	121	SENWGHFRSHHDYIYIATSDGTREGLESTQGGVCLTRKVKVILRVQSPRGCAVPPKP	180
QY	181	VSEMEMERDGAHSLFEGKENTLPDPTSNATSGABEGLPPSPVAVAGAGATALLL	240
DB	181	VSEMEMERDGAHSLFEGKENTLPDPTSNATSGABEGLPPSPVAVAGAGATALLL	240

QY 241 GYAGAGGACWRRRRRAKPSRHPGSGTGLGGGGMGPRAEPGELIALRG 300  
DB 241 GYAGAGGACWRRRRRAKPSRHPGSGTGLGGGGMGPRAEPGELIALRG 300  
QY 301 GAADPPFCPHYKVSQDYGHPYIVODGPPQSPNITYKV 340  
DB 301 GAADPPFCPHYKVSQDYGHPYIVODGPPQSPNITYKV 340

## RESULT 2

US-08-436-044-2  
Sequence 2, Application US/08436044  
Patent No. 5624899  
GENERAL INFORMATION:  
APPLICANT: Bennett, Brian D.  
APPLICANT: Matthews, William  
TITLE OF INVENTION: HTK LIGAND  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/436.044  
FILING DATE: 05-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/277722  
FILING DATE: 20-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 00,000  
REFERENCE/DOCKET NUMBER: 902D3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1994  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 336 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-436-044-2

Query Match 34.5%; Score 637.5; DB 1; Length 336;  
Best Local Similarity 42.1%; Pred. No. 1e-45;  
Matches 139; Conservative 49; Mismatches 129; Indels 13; Gaps 5;  
QY 14 GALLILGVLGSLSEFVYVNSANKRFOAGGYLYPOIGDRDLICPRARPPGPHSS 73  
DB 17 GLMLVLCRTAISRISYLEPIYVNSNSKFLPGQGLVLYPOIGDKDIIICPKV--DSKTV 73  
QY 74 PNYEFKLVVGAAGRCRCEAPPAPNLLITCDPDLRFTIKFOYSFNLGHEFRSHH 133  
DB 74 GQYEVYVYVWDKQADRCITIKENTPILNCARPDQDVFTIKFOEFSNLMGLEFOQK 133  
QY 134 DYTIIATSDGTREGLSLGGVCLTRGMVLLRVGO--SPRGAVPRKPVSEMPMER-DR 190  
DB 134 DYTIIATSDGTREGLSLGGVCLTRGMVLLRVGO--SPRGAVPRKPVSEMPMER-DR 190  
QY 191 GAHSLPEKXENLPGDPTSNATSRGAEGPLPPSPMVAVGAAGLALLLGVAGAGAMC 250  
DB 194 GRSTTSPFKVPKPGSGTSDNSAGHSGNNLGSVALFAGIASGCIIFVITITVLL 253

QY 251 WRRRRRAKPSRHPGSGTGLGGGGMGPRAEPGELIALRGGAADPPFCPH 310  
DB 254 KYRRRRRAKPSRHPGSGTGLGGGGMGPRAEPGELIALRGGAADPPFCPH 306  
QY 311 YKVSQDYGHPYIVODGPPQSPNITYKV 340  
DB 307 YKVSQDYGHPYIVODGPPQSPNITYKV 336

## RESULT 3

US-08-436-054-2  
Sequence 2, Application US/08436054  
Patent No. 5864020  
GENERAL INFORMATION:  
APPLICANT: Bennett, Brian D.  
APPLICANT: Matthews, William  
TITLE OF INVENTION: HTK LIGAND  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/436.054  
FILING DATE: 05-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/277722  
FILING DATE: 20-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 00,000  
REFERENCE/DOCKET NUMBER: 902D1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1994  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 336 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-436-054-2

Query Match 34.5%; Score 637.5; DB 2; Length 336;  
Best Local Similarity 42.1%; Pred. No. 1e-45;  
Matches 139; Conservative 49; Mismatches 129; Indels 13; Gaps 5;  
QY 14 GALLILGVLGSLSEFVYVNSANKRFOAGGYLYPOIGDRDLICPRARPPGPHSS 73  
DB 17 GLMLVLCRTAISRISYLEPIYVNSNSKFLPGQGLVLYPOIGDKDIIICPKV--DSKTV 73  
QY 74 PNYEFKLVVGAAGRCRCEAPPAPNLLITCDPDLRFTIKFOYSFNLGHEFRSHH 133  
DB 74 GQYEVYVYVWDKQADRCITIKENTPILNCARPDQDVFTIKFOEFSNLMGLEFOQK 133  
QY 134 DYTIIATSDGTREGLSLGGVCLTRGMVLLRVGO--SPRGAVPRKPVSEMPMER-DR 190  
DB 134 DYTIIATSDGTREGLSLGGVCLTRGMVLLRVGO--SPRGAVPRKPVSEMPMER-DR 190  
QY 191 GAHSLPEKXENLPGDPTSNATSRGAEGPLPPSPMVAVGAAGLALLLGVAGAGAMC 250  
DB 194 GRSTTSPFKVPKPGSGTSDNSAGHSGNNLGSVALFAGIASGCIIFVITITVLL 253  
QY 251 WRRRRRAKPSRHPGSGTGLGGGGMGPRAEPGELIALRGGAADPPFCPH 310

Db 254 KYRRHRHKSPOHTTTLSTLATPRKGN-----NSERSDVIPLR---TDSVFCGH 306  
 QY 311 YEKVSGDYGHPIYVQDGPSPNNIYKV 340  
 Db 307 YEKVSGDYGHPIYVQDGPSPNNIYKV 336

RESULT 4  
 PCT-US95-08812-2  
 ; Sequence 2, Application PC/US9508812  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Genentech, Inc.  
 ; TITLE OF INVENTION: HTX LIGAND  
 ; NUMBER OF SEQUENCES: 7  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genentech, Inc.  
 ; STREET: 460 Point San Bruno Blvd  
 ; CITY: South San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94080  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: patin (Genentech)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US95/08812  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Lee, Wendy M.  
 ; REGISTRATION NUMBER: 00,000  
 ; REFERENCE/DOCKET NUMBER: 902PCT  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415/225-1994  
 ; TELEFAX: 415/952-9881  
 ; TELEX: 910/371-7168  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 336 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; PCT-US95-08812-2

Query Match 34.5%; Score 637.5; DB 5; Length 336;  
 Best Local Similarity 42.1%; Pred. No. 1e-45;  
 Matches 139; Conservative 49; Mismatches 129; Indels 13; Gaps 5;  
 QY 14 GALLIAGLVGLVSGLSLEPYVWNSANKRFQAGGYLYPOIGDLIDLCPRAPPGPHSS 73  
 Db 17 GLVAVLCRTAISRIIVLEPIYVWNSNSKFLBGQGLVLYPOIGDLIDLICKV---DSKIV 73  
 QY 74 PNYEFYKLYLVGAQGRRCAPAPNLLTCDRPDLRLFTIKQOYSEPLMKHEPFSHA 133  
 Db 74 GQVEYKLYVVDKQADRCIKENTPLNCAAPDQVKKFIKQOFSPLVWGLPEFQNK 133  
 QY 134 DYVLIATSDGRREGLESLOGVCLTRGKMYLRLVGO--SPRGAVPRKPPSEMMER-DR 150  
 Db 134 DYVLIATSDGRREGLESLOGVCLTRGKMYLRLVGO--SPRGAVPRKPPSEMMER-DR 150  
 QY 191 GAASLEBKENTLPDPTSNATSRGAEPLPPSPMPAVAGAAGLALLLLGVAAGAGMC 250  
 Db 194 GRSSFTSPFVXPNGSSHDGASAGHSGNNLLGSEVALFAGIASGCIIFIVIIITLVLL 253  
 QY 251 WRRRAKSESERHPPGFGFGGSLGCGGGGMPRBAPELGLIARCGGAADPPRCPH 310  
 Db 254 KYRRHRHKSPOHTTTLSTLATPRKGN-----NSERSDVIPLR---TDSVFCGH 306

QY 311 YEKVSGDYGHPIYVQDGPSPNNIYKV 340  
 Db 307 YEKVSGDYGHPIYVQDGPSPNNIYKV 336

RESULT 5  
 US-08-213-403-2  
 ; Sequence 2, Application US/08213403  
 ; Patent No. 5512457  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lyman, Stewart D.  
 ; APPLICANT: Beckmann, M. Patricia  
 ; APPLICANT: Baum, Peter R.  
 ; APPLICANT: Carpenter, Melissa  
 ; TITLE OF INVENTION: No. 5512457el Cytokine Designated elk ligand  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Immunex Corporation  
 ; STREET: 51 University Street  
 ; CITY: Seattle  
 ; STATE: Washington  
 ; COUNTRY: USA  
 ; ZIP: 98101  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: Apple Macintosh  
 ; OPERATING SYSTEM: Apple System 7.1  
 ; SOFTWARE: Microsoft Word for Apple, Version 5.1a  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/213,403  
 ; FILING DATE: 15-MAR-1994  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/977,693  
 ; FILING DATE: 13-NOV-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Seese, Kathryn A.  
 ; REGISTRATION NUMBER: 32,172  
 ; REFERENCE/DOCKET NUMBER: 2807-A  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (206) 587-0430  
 ; TELEFAX: (206) 233-0644  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 346 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-213-403-2

Query Match 34.2%; Score 632; DB 1; Length 346;  
 Best Local Similarity 39.5%; Pred. No. 3e-45;  
 Matches 145; Conservative 48; Mismatches 116; Indels 58; Gaps 9;  
 QY 8 PGAVVGAALLLLGLVGLVSGL-----SLEPYVWNSANKRFQAGGYLYPOIGDLIDL 61  
 Db 4 PGQVWLGKVLVAVVWMAICRLATPLAKLLEFVSWSSINPKFLSGGLVITYPIGDKLII 63  
 QY 62 CPRAAPPGPHSEPNYEFYKLYLVGAQGRRCAPAPNLLTCDRPDLRLFTIKQOYSE 121  
 Db 64 CPRAAPPGPHSEPNYEFYKLYLVGAQGRRCAPAPNLLTCDRPDLRLFTIKQOYSE 118  
 QY 122 PNLWGHFPRSHHDYIIITSDTREGLESLOGVCLTRGKMYLRLVGO--SPRGAVPRKPV 181  
 Db 119 PNYMGLFPRKHDYIIITSDTREGLESLOGVCLTRGKMYLRLVGO--SPRGAVPRKPV 178  
 QY 182 SMPWERDRGAHSLP-PEKENTLPDPTSNATSRGAEPLPPSPMPAVAGAAGLA---- 236  
 Db 179 SRPSKADNTVMATQAGSGSLSDSDGKHETVQOEKSGP-----GASGSSGDPD 231  
 QY 237 -----LILLVGAAGGA-----MCWRRRAKSESERHPPGFGFGGSLGL 277  
 Db 232 GPFNSKVALFAAVGAGCVIFLLIITLVLLIKLRKRKRKTQ-----RAAALSL 282

QY 278 -----GGGGGMPREAPGELIALRGGAADPPFCPHYKESGDYGHPIYVODGPPOSP 333  
DB 283 STLASPKGSGTGATGTPSDIILPLR---TTENNYPCHYEKSGDYGHPYIYQEMPPOSP 339  
QY 334 PNITYKV 340  
DB 340 ANITYKV 346

RESULT 6  
US-08-458-077-2  
Sequence 2, Application US/08458077  
Patent No. 5627267  
GENERAL INFORMATION:  
APPLICANT: Lyman, Stewart D.  
APPLICANT: Beckmann, M. Patricia  
APPLICANT: Baum, Peter R.  
APPLICANT: Carpenter, Melissa  
TITLE OF INVENTION: No. 5627267el Cytokine Designated elx ligand  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98101

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: Apple Macintosh  
SOFTWARE: Microsoft Word for Apple, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458, 077  
FILING DATE: 01-JUN-1995  
CLASSIFICATION: 5.4  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/213,403  
FILING DATE: 15-MAR-1994  
APPLICATION NUMBER: US 07/977,693  
FILING DATE: 13-NOV-1992  
CLASSIFICATION: 5.14  
ATTORNEY/AGENT INFORMATION:  
NAME: Seese, Kathryn A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2807-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 346 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-458-077-2

Query Match 34.2%; Score 632; DB 1; Length 346;  
Best Local Similarity 39.5%; Pred. No. 3e-45; Indels 58; Gaps 9;  
Matches 145; Conservative 48; Mismatches 116;  
QY 8 PGGRVAGALLLGLVLSGL-----SLPPYVNSANKRFOAEGGYLVLPQIGDRLLD 61  
DB 4 PGGRWLGKMLVAVVWMLCRLATPLAKNLEPVSWSLNPKFLSGKGLVIYPKIGKLDII 63  
QY 62 CPAPAPPGPHSSPNYEFYLYLVGAQGRCAEPAPNILLICDDPDDDLRTITFOEYS 121  
DB 64 CPFAEAGRP-----YERYKLYVRPQAAACSTVLDENVLVTCNRPESGIRFTIKFOES 118  
QY 122 PNIMGHFRSHDYIYIATSDGTREGLESLOGVCLTRGKYLRLRVGQSPRGAVPRXPV 181  
DB 119 PNIMGLEFKKHHDYIYITSTNGSLBLENREGVCRTRMKIMKYGDPPNMTVEQLTT 178

QY 182 SEMMERDRGAASLE-PGKENLPDPTSNATSRGAEGFLPPSPMPAVAGAAAGLA----- 236  
DB 179 SRPSKADNTVKNATAPSGRSLGSDSKHETVNSEKSGP-----GASGSSGDD 231  
QY 237 -----LLILGVAGAGA-----NCWRRRRAKPSBSRHPGSGFRGSGSLG 277  
DB 232 GFENSKVALFAAVAGACVIFLLIIFLVLLKLRGRHRKHQQ-----RAAALSL 282  
QY 278 -----GGGGGMPREAPGELIALRGGAADPPFCPHYKESGDYGHPIYVODGPPOSP 333  
DB 283 STLASPKGSGTGATGTPSDIILPLR---TTENNYPCHYEKSGDYGHPYIYQEMPPOSP 339  
QY 334 PNITYKV 340  
DB 340 ANITYKV 346

RESULT 7  
US-08-460-741-2  
Sequence 2, Application US/08460741  
Patent No. 5670625  
GENERAL INFORMATION:  
APPLICANT: Lyman, Stewart D.  
APPLICANT: Beckmann, M. Patricia  
APPLICANT: Baum, Peter R.  
APPLICANT: Carpenter, Melissa  
TITLE OF INVENTION: No. 5670625el Cytokine Designated elx ligand  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98101

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: Apple Macintosh  
SOFTWARE: Microsoft Word for Apple, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/460,741  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 5.30  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/213,403  
FILING DATE: 15-MAR-1994  
APPLICATION NUMBER: US 07/977,693  
FILING DATE: 13-NOV-1992  
CLASSIFICATION: 5.30  
ATTORNEY/AGENT INFORMATION:  
NAME: Seese, Kathryn A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2807-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 346 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-460-741-2

Query Match 34.2%; Score 632; DB 1; Length 346;  
Best Local Similarity 39.5%; Pred. No. 3e-45; Indels 58; Gaps 9;  
Matches 145; Conservative 48; Mismatches 116;  
QY 8 PGGRVAGALLLGLVLSGL-----SLPPYVNSANKRFOAEGGYLVLPQIGDRLLD 61  
DB 4 PGGRWLGKMLVAVVWMLCRLATPLAKNLEPVSWSLNPKFLSGKGLVIYPKIGKLDII 63

QY 62 CPRAAPGPHSSPNVEFYKLYLVGAQGRCEAPAPNLLTCDPDLRLFTIKFOEYS 121  
 Db 64 CPRAEAGRP-----YEEYKLYLVPEQAAACSTVLDPNVLTCNRPQEIIRFTIKFOEFS 118  
 QY 122 PNLGHEFRSHDYYIIATSDGTREGLESIOGGVCLTRGKTVLLRVGQSPRGAVPRKPV 181  
 Db 119 PNYMGLBPKKHHDYIITSNSNSLBLEGNREGVCTRIMKIKMGQDPNVAVTPEQLTT 178  
 QY 182 SEMEMERDRGAHSLR-PGKENTPGDPTSNATSRGAEPLPPSPMAVAGAAGLA---- 236  
 Db 179 SRPSKADNTVYKATQAPGSRGSLGSDGKHETVNOEKSQP-----GASGSSGDPD 231  
 QY 237 -----LLLVGAAGA-----MCKRRRAKPSRHRPGPSFGSGSLG 277  
 Db 232 GFNSKVALPAVAGACVITLLIITVLLLKLRKHRTQO-----RAALSL 282  
 QY 278 ----GGGGMGPREAPFEGELIALRGGAADPPFCPHYEKVSGDYGHPIYIVODGPPSP 333  
 Db 283 STLASPKGSGTAGTEPSDIIIPLR---TTENNYCPHYEKVSGDYGHPIYIVQEMPPOSP 339  
 QY 334 ENIYKYV 340  
 Db 340 ANIYKYV 346

# RESULT 8

US-08-747-240-2  
 / Sequence 2, Application US/08747240  
 / Patent No. 5728813  
 / GENERAL INFORMATION:  
 / APPLICANT: Lyman, Stewart D.  
 / APPLICANT: Beckmann, M. Patricia  
 / APPLICANT: Baum, Peter R.  
 / APPLICANT: Carpenter, Melissa  
 / TITLE OF INVENTION: No. 5728813el Cytokine Designated alk ligand  
 / NUMBER OF SEQUENCES: 2  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: Immunex Corporation  
 / STREET: 51 University Street  
 / CITY: Seattle  
 / STATE: Washington  
 / COUNTRY: USA  
 / ZIP: 98101  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Floppy disk  
 / COMPUTER: Apple Macintosh  
 / OPERATING SYSTEM: Apple System 7.1  
 / SOFTWARE: Microsoft Word for Apple, Version 5.1a  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/747,240  
 / FILING DATE: 12-NOV-1996  
 / CLASSIFICATION: 435  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: US 08/213,403  
 / FILING DATE: 15-MAR-1994  
 / APPLICATION NUMBER: US 07/977,693  
 / FILING DATE: 13-NOV-1992  
 / CLASSIFICATION: 435  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Seese, Kathryn A.  
 / REGISTRATION NUMBER: 32,172  
 / REFERENCE/DOCKET NUMBER: 2807-A  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: (206) 587-0430  
 / TELEFAX: (206) 233-0644  
 / INFORMATION FOR SEQ ID NO: 2:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 346 amino acids  
 / TYPE: amino acid  
 / TOPOLOGY: linear  
 / MOLECULE TYPE: protein  
 / US-08-747-240-2

Query Match 34.2%; Score 632; DB 1; Length 346;  
 Best Local Similarity 39.5%; Pred. No. 3e-45;  
 Matches 145; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

QY 8 PGGRVVGALLLVGLVSGP-----SLSPVYNSANKRFOAGGYLYVQIDRDL 61  
 Db 4 PGGRWLGKVLVAMVYMLCRLATPLACNLSPVSSINPFLSGKGLVYIPKIDKDI 63  
 QY 62 CPRAAPGPHSSPNVEFYKLYLVGAQGRCEAPAPNLLTCDPDLRLFTIKFOEYS 121  
 Db 64 CPRAEAGRP-----YEEYKLYLVPEQAAACSTVLDPNVLTCNRPQEIIRFTIKFOEFS 118  
 QY 122 PNLGHEFRSHDYYIIATSDGTREGLESIOGGVCLTRGKTVLLRVGQSPRGAVPRKPV 181  
 Db 119 PNYMGLBPKKHHDYIITSNSNSLBLEGNREGVCTRIMKIKMGQDPNVAVTPEQLTT 178  
 QY 182 SEMEMERDRGAHSLR-PGKENTPGDPTSNATSRGAEPLPPSPMAVAGAAGLA---- 236  
 Db 179 SRPSKADNTVYKATQAPGSRGSLGSDGKHETVNOEKSQP-----GASGSSGDPD 231  
 QY 237 -----LLLVGAAGA-----MCKRRRAKPSRHRPGPSFGSGSLG 277  
 Db 232 GFNSKVALPAVAGACVITLLIITVLLLKLRKHRTQO-----RAALSL 282  
 QY 278 ----GGGGMGPREAPFEGELIALRGGAADPPFCPHYEKVSGDYGHPIYIVODGPPSP 333  
 Db 283 STLASPKGSGTAGTEPSDIIIPLR---TTENNYCPHYEKVSGDYGHPIYIVQEMPPOSP 339  
 QY 334 ENIYKYV 340  
 Db 340 ANIYKYV 346

# RESULT 9

US-08-239-567-6  
 / Sequence 6, Application US/08299567  
 / Patent No. 5747033  
 / GENERAL INFORMATION:  
 / APPLICANT: Davis, et al.  
 / TITLE OF INVENTION: METHOD OF ENHANCING THE BIOLOGICAL  
 / TITLE OF INVENTION: ACTIVITY OF EPH FAMILY LIGANDS  
 / NUMBER OF SEQUENCES: 8  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: Regeneron Pharmaceuticals, Inc.  
 / STREET: 777 Old Saw Mill River Road  
 / CITY: Tarrytown  
 / STATE: New York  
 / COUNTRY: U.S.A.  
 / ZIP: 10591-6707  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Floppy disk  
 / COMPUTER: IBM PC compatible  
 / OPERATING SYSTEM: PC-DOS/MS-DOS  
 / SOFTWARE: PatentIn Release #1.0, Version #1.30  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/299,567  
 / FILING DATE: 01-SEP-1994  
 / CLASSIFICATION: 435  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Kempster, Gail M.  
 / REGISTRATION NUMBER: 32,143  
 / REFERENCE/DOCKET NUMBER: REG 290  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: 914-345-7400  
 / TELEFAX: 914-345-7721  
 / INFORMATION FOR SEQ ID NO: 6:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 346 amino acids  
 / TYPE: amino acid  
 / STRANDEDNESS:  
 / TOPOLOGY: unknown  
 / MOLECULE TYPE: protein



US-08-299-567-6

Query Match 34.2%, Score 632, DB 1, Length 346;  
 Best Local Similarity 39.5%, Pred. No. 3e-45;  
 Matches 145; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

```

QY 8 PGGVVGLLLGLVGLVGL-----SLERYNNSANKRQAGGYLVLPQIDRLDL 61
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 4 PQGVVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLV
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 62 CRRAPPGRPHSSPNYEFKLYLVGAQGRCEAPPAPVLLTCRPLDLFTIKFOEYS 121
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 64 CRRAPGRP-----YEVKLYLVREPEQAAACSTYLDPRVLTQCRPEEIRFTIKFOEYS 118
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 122 PNLWGEFRSHHDYITLTSOGTREGLESLOGGVCITFGMKVLLRVGSGPRGAVPKRPV 181
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 119 PNYMLERFKHHDYITLTSNGSLGLELRBGVCRTITMKIMKVGQDPNAVAPPEQLTT 178
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 182 SEMPERDRGAHSLR-GEKENTLPGDPTSNATSRGAEGLPPSPMPAVAGAAGLA---- 236
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 179 SRPSKEDADVTVMATQAPESRGLSDSDGKHETVNGEKSGP-----GASGSSGDDPD 231
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 237 -----LLLLGVAGAGCA-----MCWRRRAKPSERHPPGSGFGRGSLGL 277
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 232 GFENSKVALLFAVAGAGCVTLFLITLFLTVLLKLRKRRTKHTQ-----RAALLSL 282
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 278 ---GGGGMGREAPPELGLALGGGAADPPCFPHYKSGDYGHVYTVQDPPQSP 333
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 283 SLIAPKSGSGTAGTEPDIIIFLR---TENNVCPEHYKSGDYGHVYTVQEMPQSP 339
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 334 PNIYKYV 340
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 340 ANIYKYV 346
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 10

US-08-436-044-4  
 ; Sequence 4, Application US/08436044  
 ; Patent No. 5624699

; GENERAL INFORMATION:

APPLICANT: Bennett, Brian D.  
 APPLICANT: Matthews, William  
 TITLE OF INVENTION: HTK LIGAND  
 NUMBER OF SEQUENCES: 7  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Genentech, Inc.  
 STREET: 460 Point San Bruno Blvd  
 CITY: South San Francisco  
 STATE: California  
 COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: patin (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/436,044

FILING DATE: 05-MAY-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/277722

FILING DATE: 20-JUL-1994

ATTORNEY/AGENT INFORMATION:

NAME: Lee, Wendy M.

REGISTRATION NUMBER: 00,000

REFERENCE/DOCKET NUMBER: 902D3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/225-1994

TELEFAX: 415/952-9881

TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 333 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-436-044-4

Query Match 34.0%, Score 629.5, DB 1, Length 333;  
 Best Local Similarity 40.9%, Pred. No. 4.6e-45;  
 Matches 135; Conservative 52; Mismatches 130; Indels 13; Gaps 5;

```

QY 14 GALLLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLV
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 14 GVLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLV
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 74 PNYEFKLYLVGAQGRCEAPPAPVLLTCRPLDLFTIKFOEYSFNLWGEFRSHH 133
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 71 GQYEVKLYVMVVKDQADRCTIKENTPLNCAKPPQDIFITIKFOEYSFNLWGEFRSHH 130
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 134 DYIITSDGTREGLESLOGGVCITFGMKVLLRVGQ--SPRGAVPRKPSSEMPER-DR 190
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 131 DYIITSDGTREGLESLOGGVCITFGMKVLLRVGQ--SPRGAVPRKPSSEMPER-DR 190
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 191 GAHSLPEKENTLPGDPTSNATSRGAEGLPPSPMPAVAGAAGLALLLVGAAGAMC 250
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 191 GRSTTSPYKPNPGSGSTIDGNSGNNILGSEVALFAGISGCIIFIVITLVLL 250
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 251 WRRRAKPSERHPPGSGRGSGLGGGGMGREAPPELGLALGGGAADPPCFPH 310
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 251 KTRRRKHSPOHTTTLSTLATPRSGN---NGSPSDIIFLR---TADSVFCH 303
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 311 YEKVSGDYGHVYTVQDPPQSPNIYKYV 340
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 304 YEKVSGDYGHVYTVQEMPQSPANIYKYV 333
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 11

US-08-436-054-4  
 ; Sequence 4, Application US/08436054  
 ; Patent No. 5864020

; GENERAL INFORMATION:

APPLICANT: Bennett, Brian D.  
 APPLICANT: Matthews, William  
 TITLE OF INVENTION: HTK LIGAND  
 NUMBER OF SEQUENCES: 7  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Genentech, Inc.  
 STREET: 460 Point San Bruno Blvd  
 CITY: South San Francisco  
 STATE: California  
 COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: patin (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/436,054

FILING DATE: 05-MAY-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/277722

FILING DATE: 20-JUL-1994

ATTORNEY/AGENT INFORMATION:

NAME: Lee, Wendy M.

REGISTRATION NUMBER: 00,000

REFERENCE/DOCKET NUMBER: 902D1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/225-1994

TELEFAX: 415/952-9881

TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 333 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
US-08-436-054-4

Query Match  
Best Local Similarity 40.9%; Score 629.5; DB 2; Length 333;  
Matches 135; Conservative 52; Mismatches 130; Indels 13; Gaps 5;

14 GALLILGVLGSLVLSLEPYWNSANKRFQAEGLVLYPQIGRLDLCPARPPEHSS 73  
14 GVLMLCRATISKSIYLEPIYWNSSKFLPGQGLVLYPQIGKDLICPKV---DSKIV 70  
74 PNYEFKLYVGAQGRCEAPAPNLLTCDRPDLRFITKFOEYSFNLNGHEFRSH 133  
71 GQYEVYKVMYVDKQADRCITKENTPLNCAKPDODIKFTIKFOEFSPLNGLEFQK 130

134 DYIITSDTREGLESLOGVCLTRGMKVLRVGQ--SPRGAVPRKVSMPMER-DR 190  
131 DYIITSDTREGLESLOGVCLTRGMKVLRVGQ--SPRGAVPRKVSMPMER-DR 190  
191 GAASLEPGENLPDPTSNATSRGAEGPLPPSPMPAVAGAGLALLLVAGAAGAMC 250  
191 GRSTTSPFKPNPQSGSTDSNAGSGNNILGSEVALFAGIASGCTIFVIITLVLL 250  
251 WRRRAKPSERHPGPGSFGGSLGCGGGMGPREAPBELGIALRGGAADPPFCPH 310  
251 KYRRHRKHSQHTTSLSTLATPKRSNN---NGSEPSDIIIPLR---TADSVFCPH 303

311 YEKVSGDGHVYIVODGPPQSPENIYKYV 340  
304 YEKVSGDGHVYIVQEMPPQSPANIYKYV 333

RESULT 12  
US-08-271-948-2  
Sequence 2, Application US/08271948  
Patent No. 6303769  
GENERAL INFORMATION:  
APPLICANT: Cellecti, Douglas P.  
APPLICANT: Reddy, Pranthitha  
TITLE OF INVENTION: No. 6303769e1 Cytokine Designated Lerk-5  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: Washington  
COUNTRY: US  
ZIP: 98101

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple 7.1  
SOFTWARE: Microsoft Word, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/271,948  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Seese, Kathryn A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2823  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
TELEX: 756822

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 333 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-271-948-2

Query Match  
Best Local Similarity 40.9%; Score 629.5; DB 4; Length 333;  
Matches 135; Conservative 52; Mismatches 130; Indels 13; Gaps 5;

14 GALLILGVLGSLVLSLEPYWNSANKRFQAEGLVLYPQIGRLDLCPARPPEHSS 73  
14 GVLMLCRATISKSIYLEPIYWNSSKFLPGQGLVLYPQIGKDLICPKV---DSKIV 70  
74 PNYEFKLYVGAQGRCEAPAPNLLTCDRPDLRFITKFOEYSFNLNGHEFRSH 133  
71 GQYEVYKVMYVDKQADRCITKENTPLNCAKPDODIKFTIKFOEFSPLNGLEFQK 130

134 DYIITSDTREGLESLOGVCLTRGMKVLRVGQ--SPRGAVPRKVSMPMER-DR 190  
131 DYIITSDTREGLESLOGVCLTRGMKVLRVGQ--SPRGAVPRKVSMPMER-DR 190  
191 GAASLEPGENLPDPTSNATSRGAEGPLPPSPMPAVAGAGLALLLVAGAAGAMC 250  
191 GRSTTSPFKPNPQSGSTDSNAGSGNNILGSEVALFAGIASGCTIFVIITLVLL 250  
251 WRRRAKPSERHPGPGSFGGSLGCGGGMGPREAPBELGIALRGGAADPPFCPH 310  
251 KYRRHRKHSQHTTSLSTLATPKRSNN---NGSEPSDIIIPLR---TADSVFCPH 303

311 YEKVSGDGHVYIVODGPPQSPENIYKYV 340  
304 YEKVSGDGHVYIVQEMPPQSPANIYKYV 333

RESULT 13  
PCT-US95-08534-2  
Sequence 2, Application PC/TUS9508534  
GENERAL INFORMATION:  
APPLICANT: Immunex Corporation  
TITLE OF INVENTION: Novel Cytokine Designated Lerk-5  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: Washington  
COUNTRY: US  
ZIP: 98101

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple 7.1  
SOFTWARE: Microsoft Word, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/08534  
FILING DATE: 06-JUL-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/271,948  
FILING DATE: 08-JUL-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Anderson, Kathryn A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2823-WO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
TELEX: 756822

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 333 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-08534-2

Query Match	34.0%;	Score 629.5;	DB 5;	Length 333;
Best Local Similarity	40.9%;	Pred. No. 4.6e-45;		
Matches 135;	Conservative 52;	Mismatches 130;	Indels 13;	Gaps 5

[illegible]

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	34.0%	Score 629.5;	DB 5;	Length 333;
Best Local Similarity	40.9%	Pred. No. 4.6e-45;		
Matches 135; Conservative	52;	Mismatches 130;	Indels 13;	Gaps 5;

[illegible]

```

RESULT 15
US-09-214-631-4
: Sequence 4, Application US/09214631
: Patent No. 6413730
: GENERAL INFORMATION:
: APPLICANT: Holland, Sacha
: APPLICANT: Mbamalu, Geraldine
: APPLICANT: Pawson, Tony
: TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
: TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
: TITLE OF INVENTION: TYROSINE KINASES
: FILE REFERENCE: 11757.23USMO
: CURRENT APPLICATION NUMBER: US/09/214,631
: CURRENT FILING DATE: 1999-03-12
: EARLIER APPLICATION NUMBER: PCT/CA97/00473
: EARLIER FILING DATE: 1997-07-04
: EARLIER APPLICATION NUMBER: 60/021,272
: EARLIER FILING DATE: 1996-07-05
: NUMBER OF SEQ ID NOS: 13
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 4
: LENGTH: 333
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-214-631-4

```

[illegible]

```

Db 251 KYRRRRKSPQHTTTLSTLSTLAPKRSKNN---NGSEPSDIIIPK---TADSVFCH 303
QY 311 YEKVSGDYGHPYIVODGPQSPNNIYKV 340
Db 304 YEKVSGDYGHPYIVODGPQSPNNIYKV 333

```

```

RESULT 16
US-09-214-631-5
; Sequence 5, Application US/09214631
; Patent No. 6413730
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawsom, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/09/214,631
; EARLIER FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: PCT/CA97/00473
; EARLIER FILING DATE: 1997-07-04
; EARLIER APPLICATION NUMBER: 60/021,272
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-631-5

```

```

Query Match 32.9%; Score 608.5; DB 4; Length 345;
Best Local Similarity 38.7%; Pred. No. 2,7e-43;
Matches 142; Conservative 50; Mismatches 116; Indels 59; Gaps 10;

```

```

QY 8 PGGVAVG---ALLLLVGLVSGL--SLPEVTWNSANKRFQABGGVYVPOIGDRIDLI 61
Db 4 PGGWLTAKWLVAMVMAICRLAPLANLEPVSMSSLNPKFELGKGIVIVPKIGDLDLI 63
QY 62 CPFRAPRPGSHSPYEFYKLYLVGAQGRCEAPRAPNLLLTCDRPLDLRFTIKFOEYS 121
Db 64 CPRLAARP-----YEVYKLYVRPQAACSTYLDPMVLTQNRBQIRFTIKFOES 118
QY 122 PNLWGEHFRSHDYIATSDGTREGLESLOGGVCLTRKMKVLLRVGQSPRGAVPRKPY 181
Db 119 PNYVGLTFKKHHDYITSTENGSLGLEENREGVCRTTKIMKYQODPNNAVTPQLTT 178
QY 182 SEMMERDRGAASLE-PGKENLPGDPTSNATSRGAEPLRPSPMAVVGAAAGLA----- 236
Db 179 SRPEKEDNTVYKATQAPGRBSGLGSDGKHETVNOEKGCP-----GASGSSGDDP 231
QY 237 -----LLLLGVAGAGA-----MCVRRRAKPSRSHRPGPSFGRGSLGL 277
Db 232 GFENSKVALFAAVAGACVILLITIPVTLTKLPKXKRGHTQ-----RAALSL 281
QY 278 -----GGGGGMPREAREGEIGIALRGGAADPPCCPYEKVSGDYGHPYIVODGPQSP 333
Db 282 STIASPGSGSTAGTSPDIIIPL---FTTENYCPHYEKVSGDYGHPYIVODGPQSP 338
QY 334 PNIYKV 340
Db 339 ANIYKV 345

```

```

RESULT 17
US-09-214-631-13
; Sequence 13, Application US/09214631
; Patent No. 6413730
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha

```

```

; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawsom, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/09/214,631
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: PCT/CA97/00473
; EARLIER FILING DATE: 1997-07-04
; EARLIER APPLICATION NUMBER: 60/021,272
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-631-13

```

```

Query Match 26.9%; Score 498; DB 4; Length 89;
Best Local Similarity 100.0%; Pred. No. 8.6e-35;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 252 RRRRAKPSRSHRPGSGRGSGLGCGGGMGPREAREPELIGIALRGGAADPPFCBY 311
Db 1 RRRRAKPSRSHRPGSGRGSGLGCGGGMGPREAREPELIGIALRGGAADPPFCBY 60
QY 312 EKVSQDYGHPYIVODGPQSPNNIYKV 340
Db 61 EKVSQDYGHPYIVODGPQSPNNIYKV 89

```

```

RESULT 18
US-09-214-631-11
; Sequence 11, Application US/09214631
; Patent No. 6413730
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mbamalu, Geraldine
; APPLICANT: Pawsom, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/09/214,631
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: PCT/CA97/00473
; EARLIER FILING DATE: 1997-07-04
; EARLIER APPLICATION NUMBER: 60/021,272
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-631-11

```

```

Query Match 10.9%; Score 201; DB 4; Length 82;
Best Local Similarity 47.3%; Pred. No. 5.1e-10;
Matches 44; Conservative 9; Mismatches 24; Indels 16; Gaps 3;

```

```

QY 252 RRRRAKPSRSHRPGSGRGSGLGCGGGMGPREAREPELIGIALRGGAADPPF 307
Db 2 RRRRAKPSRSHRPGSGRGSGLGCGGGMGPREAREPELIGIALRGGAADPPF 307
QY 308 CPHYEKVSGDYGHPYIVODGPQSPNNIYKV 340
Db 50 CPHYEKVSGDYGHPYIVODGPQSPNNIYKV 82

```

```

RESULT 19

```

US-09-214-631-12  
Sequence 12, Application US/09214631  
Patent No. 6413730  
GENERAL INFORMATION:  
APPLICANT: Holland, Sacha  
APPLICANT: Mbamalu, Geraldine  
APPLICANT: Pawson, Tony  
TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED  
TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR  
TITLE OF INVENTION: TYROSINE KINASES  
FILE REFERENCE: 11757.23USWO  
CURRENT FILING DATE: US/09/214,631  
EARLIER APPLICATION NUMBER: PCT/CA97/00473  
EARLIER FILING DATE: 1997-07-04  
EARLIER APPLICATION NUMBER: 60/021,272  
EARLIER FILING DATE: 1996-07-05  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 12  
LENGTH: 82  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-214-631-12

Query Match 10.8%; Score 200.5; DB 4; Length 82;  
Best Local Similarity 71.7%; Pred. No. 5,6e-10;  
Matches 38; Conservative 4; Mismatches 8; Indels 3; Gaps 1;

QY 288 ARBGEIGLRGGAGADPPFCPEYKVSQDGHAPYIVIQDGPSPPIYKYV 340  
DB 33 SEPSDIIPLR--TADSVCFPEYKVSQDGHAPYIVQEMPPSPPIYKYV 82

RESULT 20  
US-08-299-567-5  
Sequence 5, Application US/08299567  
Patent No. 5747033  
GENERAL INFORMATION:  
APPLICANT: Davis, et al.  
TITLE OF INVENTION: METHOD OF ENHANCING THE BIOLOGICAL  
TITLE OF INVENTION: ACTIVITY OF EPH FAMILY LIGANDS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Regeneron Pharmaceuticals, Inc.  
STREET: 777 Old Saw Mill River Road  
CITY: Tarrytown  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10591-6707  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/299,567  
FILING DATE: 01-SEP-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kempler, Gail M.  
REGISTRATION NUMBER: 32,143  
REFERENCE/DOCKET NUMBER: REG 290  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 914-345-7400  
TELEFAX: 914-345-7721  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 234 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein

US-08-299-567-5

Query Match 9.7%; Score 179; DB 1; Length 234;  
Best Local Similarity 28.4%; Pred. No. 1.3e-07;  
Matches 64; Conservative 22; Mismatches 83; Indels 56; Gaps 11;

QY 7 GPGGVVALLLLIGVLGVSLSEPVYNSANKRPQAEQGYVLYPQIDRLDLCF--R 64  
DB 24 GPG-----GALG-----NRHVVYVNSSNCHLRRE-GYVQVNVVDYLDIYCPHYN 67  
QY 65 ARPPGHSSPNVEFYKLYLVGAGQRCAPAPNILLTCDBPDL---DLRFIFQEQYS 121  
DB 68 SSGAGPQGGAEQVLYLVNSKNGRTCNASQGFK-RMECNRPAPHSPIKTSKTFORS 126  
QY 122 PNLWGEFRSHDYIIATSDGTREGLESLOGGVCLTRGMKYLRLVGSQPRGAVPRKPV 181  
DB 127 APLGVEFPHAGSHYIISPTNLH-----WKCLR--MKVFCASATSHSG---EKEV 174  
QY 182 SEPP-----MERDRGAHSL-----RGKENLP 204  
DB 175 PTLPGFMGNKINVLDPGEENPQVPLEKSIQSTSPKRLHP 219

RESULT 21

US-08-240-124-2

Sequence 2, Application US/08240124  
Patent No. 5516658  
GENERAL INFORMATION:  
APPLICANT: BECKMAN, M. P.  
APPLICANT: CERRETTI, DOUGLAS P.  
TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE  
TITLE OF INVENTION: RECEPTOR HEX  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMUNEX CORPORATION  
STREET: 51 UNIVERSITY STREET  
CITY: SEATTLE  
STATE: WASHINGTON  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple System 7.1  
SOFTWARE: Microsoft Word for Apple, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/240,124  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,132  
FILING DATE: 03-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/114,426  
FILING DATE: 30-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/109,745  
FILING DATE: 20-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SEESE, KATHRYN A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2814-C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
TELEX: 756822  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 238 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-240-124-2

Query Match 9.7%; Score 179; DB 1; Length 238;  
Best Local Similarity 28.4%; Pred. No. 1.3e-07;  
Matches 65; Conservative 24; Mismatches 80; Indels 60; Gaps 12;

QY 7 GPGGVRGALLLLGLVGLSLGSLPYWNSANKRFOAEGGYLYPOIGRDLCLP--R 64  
DB 24 GPG-----GALG-----NRHAYWNSNGHLRRE-GYTVQVNVNDYLDIYCPHYN 67  
QY 65 ARPGPHSP-----NYEFYLYVGAQGRCEAPAPNLLITCDPDL---DLRFTIKF 117  
DB 68 SSGVGPAGPGGAGGAGQYLYWVNSRNGYRVCNASQGRF-RMECNRPAPHSPIKFSKFP 126  
QY 118 QEYSPNLWGHEFRSHDYIATSDGTREGLESLOGGVCLTRGMKYLIRVQSPRGAVP 177  
DB 127 QRSAPSLGYEFHAGHEHYIISTPTNLH-----WKCLR-MKVYVCCASTSHSG--- 174  
QY 178 RKPVSMP-----MERDGAASLE-----PKENLP 204  
DB 175 EKPYPTLPQFTMGPNVKINVLDEFGENPQVPLKESISGTSFKRHLF 223

## RESULT 22

US-08-453-943-2  
Sequence 2, Application US/08453943  
Patent No. 5738844  
GENERAL INFORMATION:  
APPLICANT: BECKMANN, M. P.  
TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE  
TITLE OF INVENTION: RECEPTOR HEK  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMUNEX CORPORATION  
STREET: 51 UNIVERSITY STREET  
CITY: SEATTLE  
STATE: WASHINGTON  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple System 7.1  
SOFTWARE: Microsoft Word for Apple, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,943  
FILING DATE: 30-MAY-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/240,124  
FILING DATE: 09-MAY-1994  
APPLICATION NUMBER: US 08/161,132  
FILING DATE: 03-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/114,426  
FILING DATE: 30-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/109,745  
FILING DATE: 20-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SEESE, KATHRYN A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2814-C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
TELEX: 756822  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 238 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

## US-08-453-943-2

Query Match 9.7%; Score 179; DB 1; Length 238;  
Best Local Similarity 28.4%; Pred. No. 1.3e-07;  
Matches 65; Conservative 24; Mismatches 80; Indels 60; Gaps 12;

QY 7 GPGGVRGALLLLGLVGLSLGSLPYWNSANKRFOAEGGYLYPOIGRDLCLP--R 64  
DB 24 GPG-----GALG-----NRHAYWNSNGHLRRE-GYTVQVNVNDYLDIYCPHYN 67  
QY 65 ARPGPHSP-----NYEFYLYVGAQGRCEAPAPNLLITCDPDL---DLRFTIKF 117  
DB 68 SSGVGPAGPGGAGGAGQYLYWVNSRNGYRVCNASQGRF-RMECNRPAPHSPIKFSKFP 126  
QY 118 QEYSPNLWGHEFRSHDYIATSDGTREGLESLOGGVCLTRGMKYLIRVQSPRGAVP 177  
DB 127 QRSAPSLGYEFHAGHEHYIISTPTNLH-----WKCLR-MKVYVCCASTSHSG--- 174  
QY 178 RKPVSMP-----MERDGAASLE-----PKENLP 204  
DB 175 EKPYPTLPQFTMGPNVKINVLDEFGENPQVPLKESISGTSFKRHLF 223

## RESULT 23

US-09-057-121-2  
Sequence 2, Application US/09057121  
Patent No. 5969110  
GENERAL INFORMATION:  
APPLICANT: BECKMANN, M. P.  
TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE  
TITLE OF INVENTION: RECEPTOR HEK  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMUNEX CORPORATION  
STREET: 51 UNIVERSITY STREET  
CITY: SEATTLE  
STATE: WASHINGTON  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple System 7.1  
SOFTWARE: Microsoft Word for Apple, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/057,121  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/240,124  
FILING DATE:  
APPLICATION NUMBER: US 08/161,132  
FILING DATE: 03-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/114,426  
FILING DATE: 30-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/109,745  
FILING DATE: 20-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SEESE, KATHRYN A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2814-C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
TELEX: 756822  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 238 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-09-057-121-2

Query Match 9.7%; Score 179; DB 2; Length 238;  
Best Local Similarity 28.4%; Pred. No. 1.3e-07;  
Matches 65; Conservative 24; Mismatches 80; Indels 60; Gaps 12;

QY 7 GPGGVVAGALLILGVIGVLSGLEPVYNSANKRFQABGQVLYPQIGRIDLILCP--R 64  
DB 24 GPG-----GALG-----NRHAVYNNSSNOHLRRE-GYTVQVNVNDYLDIYCPHYN 67  
QY 65 ARPPGPHSSP-----NYEFKLYLVGAQGRCEAPAPNLLITCDRPLD---DLRFITKF 117  
DB 68 SSGVGGAGPFGGAGAQVLYLVNSRNGYRTCNASQGRF-RMECNRPAPHPSPITKSEKF 126  
QY 118 QEYSPNLWGEFRSHDYIIATSDTREGLSLQGVCLTRGMKVLRLVQGSFPGCAVP 177  
DB 127 QRYSAFSLGYEFHAGHEYYIISTPTNHL-----WKCLR--MKVFCASSTHSG--- 174  
QY 178 RKPVSMP-----MERDRGAHSL-----PKENLP 204  
DB 175 EKVPPTLPQFTMGPNVKINVLDFEGENPQVKLEKSI.GTSPKREHLP 223

RESULT 24  
US-09-358-734-2

Sequence 2, Application US/09358734  
Patent No. 6274117  
GENERAL INFORMATION:  
APPLICANT: BECKMANN, M. P.  
APPLICANT: CERRETTI, DOUGLAS P.  
TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE  
TITLE OF INVENTION: RECEPTOR HEK  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMUNEX CORPORATION  
STREET: 51 UNIVERSITY STREET  
CITY: SEATTLE  
STATE: WASHINGTON  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple System 7.1  
SOFTWARE: Microsoft Word for Apple, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/358,734  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/240,124  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/114,426  
FILING DATE: 30-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/109,745  
FILING DATE: 20-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SESE, KATHRYN A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2814-C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
TELEX: 756822  
INFORMATION FOR SEQ. ID NO. 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 238 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-09-358-734-2

Query Match 9.7%; Score 179; DB 4; Length 238;  
Best Local Similarity 28.4%; Pred. No. 1.3e-07;  
Matches 65; Conservative 24; Mismatches 80; Indels 60; Gaps 12;

QY 7 GPGGVVAGALLILGVIGVLSGLEPVYNSANKRFQABGQVLYPQIGRIDLILCP--R 64  
DB 24 GPG-----GALG-----NRHAVYNNSSNOHLRRE-GYTVQVNVNDYLDIYCPHYN 67  
QY 65 ARPPGPHSSP-----NYEFKLYLVGAQGRCEAPAPNLLITCDRPLD---DLRFITKF 117  
DB 68 SSGVGGAGPFGGAGAQVLYLVNSRNGYRTCNASQGRF-RMECNRPAPHPSPITKSEKF 126  
QY 118 QEYSPNLWGEFRSHDYIIATSDTREGLSLQGVCLTRGMKVLRLVQGSFPGCAVP 177  
DB 127 QRYSAFSLGYEFHAGHEYYIISTPTNHL-----WKCLR--MKVFCASSTHSG--- 174  
QY 178 RKPVSMP-----MERDRGAHSL-----PKENLP 204  
DB 175 EKVPPTLPQFTMGPNVKINVLDFEGENPQVKLEKSI.GTSPKREHLP 223

RESULT 25

US-08-299-567-7  
Sequence 7, Application US/08299567  
Patent No. 5747033  
GENERAL INFORMATION:  
APPLICANT: Davis, et al.  
TITLE OF INVENTION: METHOD OF ENHANCING THE BIOLOGICAL  
TITLE OF INVENTION: ACTIVITY OF EPH FAMILY LIGANDS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Regeneron Pharmaceuticals, Inc.  
STREET: 777 Old Saw Mill River Road  
CITY: Tarrytown  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10591-6707  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/299,567  
FILING DATE: 01-SEP-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kempler, Gail M.  
REGISTRATION NUMBER: 32,143  
REFERENCE/DOCKET NUMBER: REG 290  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 914-345-7400  
TELEFAX: 914-345-7721  
INFORMATION FOR SEQ. ID NO. 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-299-567-7  
Query Match 9.5%; Score 176.5; DB 1; Length 135;  
Best Local Similarity 31.9%; Pred. No. 1.1e-07;  
Matches 43; Conservative 14; Mismatches 73; Indels 5; Gaps 2;

QY 33 VYNSANKRFQABGQVLYPQIGRIDLILCP--RARPFGHSSPNYEFKLYLVGAQGR 90  
DB 1 VYNSNPKFXRREGYITVYXNDYLDILCPHYEXXXXXXXGXXECXLYLVXXEXYX 60  
QY 91 RCEAPAPNLLITCDRP---DLDLRFITKQYEVNLMGHEFRSHDYIIATSDTREG 147





TELEFAX: (206) 233-0644  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 184 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-173-492-2

Query Match 9.5%; Score 176; DB 4; Length 184;  
Best Local Similarity 29.3%; Pred. No. 1.7e-07;  
Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;

QY 33 VYVNSANKRFOAE-----GGYLYPQIGRDLDCPRAPPQPHSSPNVEFYKLYVGGA 87  
DB 10 VYVNSNPRFQYSAVGGGTYVEVSINDYDLYCPHYGAPLP-PAEMERYILVMNGE 68  
QY 88 QGRCEAPPAPNLLTCDPDL---DLRFTIKQEYSPNLWGHEFRSHDYIITSDGT 144  
DB 69 GHASCDHQRGKFWKWCNRPAPAGPLKFSEKFLFTFPSLGFERRPGHEYYIISATP-- 126  
QY 145 REGESLGGGCVLTTRGMKYLIVGQSPRGAVPRKPVSEMERDRGAHSLPEKKNLP 204  
DB 127 ---PMLVDRPCLR--LKYYVR-----FTNETLY 149

QY 205 GDP---TSNATSRGAEG 218  
DB 150 EAPRPFTSNSSCSGLGG 167

RESULT 29  
US-09-173-133-2  
; Sequence 2, Application US/09173133  
; Patent No. 6232447

; GENERAL INFORMATION:  
; APPLICANT: Cerretti, Douglas P.  
; TITLE OF INVENTION: Cytokine Designated LERK-6  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Power Macintosh  
; OPERATING SYSTEM: System 7.6  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/173,133  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/920,440  
; FILING DATE: 29-AUG-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Henry, Janis C.  
; REGISTRATION NUMBER: 34,347  
; REFERENCE/DOCKET NUMBER: 2826-B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 587-0430  
; TELEFAX: (206) 233-0644  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 184 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-173-133-2

Query Match 9.5%; Score 176; DB 4; Length 184;

Best Local Similarity 29.3%; Pred. No. 1.7e-07;  
Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;

QY 33 VYVNSANKRFOAE-----GGYLYPQIGRDLDCPRAPPQPHSSPNVEFYKLYVGGA 87  
DB 10 VYVNSNPRFQYSAVGGGTYVEVSINDYDLYCPHYGAPLP-PAEMERYILVMNGE 68  
QY 88 QGRCEAPPAPNLLTCDPDL---DLRFTIKQEYSPNLWGHEFRSHDYIITSDGT 144  
DB 69 GHASCDHQRGKFWKWCNRPAPAGPLKFSEKFLFTFPSLGFERRPGHEYYIISATP-- 126  
QY 145 REGESLGGGCVLTTRGMKYLIVGQSPRGAVPRKPVSEMERDRGAHSLPEKKNLP 204  
DB 127 ---PMLVDRPCLR--LKYYVR-----FTNETLY 149

RESULT 30  
US-09-165-533-2  
; Sequence 2, Application US/09165533  
; Patent No. 6268482

; GENERAL INFORMATION:  
; APPLICANT: Cerretti, Douglas P.  
; TITLE OF INVENTION: Cytokine Designated as LERK-6  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: System 7.1  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/165,533  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/538,709  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Malaska, Stephen L.  
; REGISTRATION NUMBER: 32,655  
; REFERENCE/DOCKET NUMBER: 2826  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 587-0430  
; TELEFAX: (206) 233-0644  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 184 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-165-533-2

Query Match 9.5%; Score 176; DB 4; Length 184;  
Best Local Similarity 29.3%; Pred. No. 1.7e-07;  
Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;

QY 33 VYVNSANKRFOAE-----GGYLYPQIGRDLDCPRAPPQPHSSPNVEFYKLYVGGA 87  
DB 10 VYVNSNPRFQYSAVGGGTYVEVSINDYDLYCPHYGAPLP-PAEMERYILVMNGE 68  
QY 88 QGRCEAPPAPNLLTCDPDL---DLRFTIKQEYSPNLWGHEFRSHDYIITSDGT 144  
DB 69 GHASCDHQRGKFWKWCNRPAPAGPLKFSEKFLFTFPSLGFERRPGHEYYIISATP-- 126

```

CY 145 REGJESI:GGGVCILTRGKKVLLRFGQSPRGGAVPKRVSEMERDRGAASLEFGKCNLP 204
Db 127 -----PNLVDRPCLR--LKYVYR-----PTNRTLY 149
CY 205 GDP---TSNATSRGAEG 218
Db 150 EAPPIFTSNSSCSGLG 167

RESULT 31
PCT-US95-12779-2
Sequence 2, Application PC/TUS9512779
GENERAL INFORMATION:
APPLICANT: Cerretti, Douglas P.
TITLE OF INVENTION: Cytokine Designated LERK-6
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.1
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/12779
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Malaika, Stephen L.
REGISTRATION NUMBER: 32,655
REFERENCE/DOCKET NUMBER: 2826-WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 233-0643
TELEFAX: (206) 587-0430
INFORMATION FOR SEQ. ID NO. 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 184 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-12779-2

Query Match 9.5%; Score 176; DB 5; Length 184;
Best Local Similarity 29.3%; Pred. No. 1.7e-07;
Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7

CY 33 VYNNASANKRFOAE-----GGYVYVYVPOIGRDLTLCPRARPPGHSSPYVEEYKLYLVGA 87
Db 10 VYNNRNSPRQVAVSANGGGYTVFVNSINDYLDYCHNGAPLP-PAEMEXYILYMNNGE 68
CY 88 QGRGCEAPAPNLLITCDPDL--DLRTTFROEYSPNLWGHFERSHADYITATSDGT 144
Db 69 GHASCIDHRQGFQKMECNRPAPAGBGLKFSERFOLFTPSLGFEPFRPCHHEYIYISATP-- 126
CY 145 REGJESI:GGGVCILTRGKKVLLRFGQSPRGGAVPKRVSEMERDRGAASLEFGKCNLP 204
Db 127 -----PNLVDRPCLR--LKYVYR-----PTNRTLY 149
CY 205 GDP---TSNATSRGAEG 218
Db 150 EAPPIFTSNSSCSGLG 167

RESULT 32
PCT-US95-15781-2
Sequence 2, Application PC/TUS9515781
GENERAL INFORMATION:

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	9.5%;	Score 176;	DB 5;	Length 209
Best Local Similarity	29.3%;	Pred. No. 2e-07;		



TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-173-492-10

Query Match 9.5%; Score 175.5; DB 4; Length 213;  
Best Local Similarity 36.8%; Pred. No. 2.3e-07;  
Matches 43; Conservative 14; Mismatches 51; Indels 9; Gaps 3;

QY 33 YVWNSANKRFOA-----EGGYLYPQIGDRDLDCPRAPPGSHSSPNVEFYKLYLVGA 87  
DB 39 YVWNSNRPFHAGAGDDGGYTVESINDYLDICPHYGALP-PAERMEHYLVYVNGE 97  
QY 88 QGRCEAPAPRLTLTCRBDL---DLAFTIKFOEYSNLMGHEFRSHDDYIATS 141  
DB 98 GHASCDHRQGRKWEKCNRPAPGGLPFSEKFLTFPSLGFEFRPGHEIYIAT 154

## RESULT 39

US-09-173-133-10  
Sequence 10; Application US/09173133

Patent No. 6232447  
GENERAL INFORMATION:

APPLICANT: Ceretti, Douglas P.  
TITLE OF INVENTION: Cytokine Designated IERK-6

NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Immunex Corporation  
STREET: 51 University Street

CITY: Seattle  
STATE: WA

COUNTRY: USA  
ZIP: 98101

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk

OPERATING SYSTEM: System 7.6  
SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/173.133

FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/920.440

FILING DATE: 29-AUG-1997  
ATTORNEY/AGENT INFORMATION:

NAME: Henry, Janis C.  
REGISTRATION NUMBER: 34,347

REFERENCE/DOCKET NUMBER: 2826-B  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644

INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:

LENGTH: 213 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein

US-09-173-133-10

Query Match 9.5%; Score 175.5; DB 4; Length 213;  
Best Local Similarity 36.8%; Pred. No. 2.3e-07;  
Matches 43; Conservative 14; Mismatches 51; Indels 9; Gaps 3;

QY 33 YVWNSANKRFOA-----EGGYLYPQIGDRDLDCPRAPPGSHSSPNVEFYKLYLVGA 87  
DB 39 YVWNSNRPFHAGAGDDGGYTVESINDYLDICPHYGALP-PAERMEHYLVYVNGE 97  
QY 88 QGRCEAPAPRLTLTCRBDL---DLAFTIKFOEYSNLMGHEFRSHDDYIATS 141  
DB 98 GHASCDHRQGRKWEKCNRPAPGGLPFSEKFLTFPSLGFEFRPGHEIYIAT 154

## RESULT 40

US-09-214-631-7

Sequence 7; Application US/09214631

Patent No. 6413730  
GENERAL INFORMATION:

APPLICANT: Holland, Sacha  
APPLICANT: Mbamalu, Geraldine

APPLICANT: Pawson, Tony  
TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED

TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR

FILE REFERENCE: 11757.23USWO  
CURRENT APPLICATION NUMBER: US/09/214.631

EARLIER APPLICATION NUMBER: PCT/CA97/00473  
EARLIER FILING DATE: 1997-07-04

EARLIER APPLICATION NUMBER: 60/021.272  
EARLIER FILING DATE: 1996-07-05

NUMBER OF SEQ ID NOS: 13  
SOFTWARE: Patent In Ver. 2.0

SEQ ID NO 7  
LENGTH: 233  
TYPE: PRT

ORGANISM: Homo sapiens  
US-09-214-631-7

Query Match 9.4%; Score 174.5; DB 4; Length 233;  
Best Local Similarity 27.4%; Pred. No. 3.1e-07;  
Matches 60; Conservative 24; Mismatches 74; Indels 61; Gaps 11;

QY 20 GVILVSGSLPVPWNSANKRFOAEGGYLYPQIGDRDLDCPR---ARPGSHSSPN 75  
DB 27 GALG-----NRHAVVWSSNOHLRE-GYTVQVWVNDYLDICPHYNSGAGPGGGAE 80  
QY 76 YEFYKLYLVG-----GAQGR---CEAPAPRLTLTCRBDLDTFTIKFOEYSNLMWGH 127  
DB 81 YVWNSNRGRTCNASQGFKEKWCNRPAPH-----SPITSEKFOYSAFSLGY 131  
QY 128 EFRSHDDYIATSDTRGLESLSQGVCLTRGMKYLTVGQSPRGAVFRKPVSEMP-- 185  
DB 132 EPHAGHEYYIYSPTHNLH-----MKCLR--MKVFCVCASTSHSG--EKPVTLPGF 179  
QY 186 -----MERDRGAHSLF-----FGKENLP 204  
DB 180 TWGPVKNVLEDFEGENPOVPLEKXISGTSPKXELHP 218

## RESULT 41

US-08-455-001-4  
Sequence 4; Application US/08455001

Patent No. 5795734  
GENERAL INFORMATION:

APPLICANT: Flanagan, John G.  
APPLICANT: Cheng, Hwai-Jong

TITLE OF INVENTION: Eph Receptor Ligands, and Uses Related  
TITLE OF INVENTION: Thereto

NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street

CITY: Boston  
STATE: MA

COUNTRY: USA  
ZIP: 02109

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII (text)

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,001

FILING DATE: 31 MAY 1995  
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

US-09-214-631-7

NAME: Vincent, Matthew P.  
 REGISTRATION NUMBER: 36,709  
 REFERENCE/DOCKET NUMBER: HMI-011CP2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 227-7400  
 TELEFAX: (617) 227-5941  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 200 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-455-001-4

Query Match 9.3%; Score 172; DB 1; Length 200;  
 Best Local Similarity 35.3%; Pred. No. 4,1e-07;  
 Matches 49; Conservative 16; Mismatches 58; Indels 16; Gaps 5;

QY 15 ALLLGLVGLVSGLSLEP-----VYVNSANKRFOAGSGYVLPQIGRLDLCPR 65  
 DB 7 AALLLAIVG-VCVMSDDPKVTSDRYAVYVNSNPRFH-RGDYTVESINDYLDYCPHY 64  
 QY 66 RPPGHSNNVEFYKLYLVGAQGRCEAPAPNLLTCDRPLD---DLRFTIKFOEYSP 122  
 DB 65 EEPFLP--AERMERVLYVMVNYEGHASCDBRQKGFKEWECNRPDSGPKFSEKQLFTF 122  
 QY 123 NMGHEFRSHDYIYIATS 141  
 DB 123 FSLGFEFRPGHEHYIYISAS 141

RESULT 42  
 PCT-US95-11869-4  
 Sequence 4, Application PC/TUS9511869  
 GENERAL INFORMATION:  
 APPLICANT:  
 TITLE OF INVENTION: Eph Receptor Ligands, and Uses Related  
 TITLE OF INVENTION: Thereeto  
 NUMBER OF SEQUENCES: 5  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE:  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US95/11869  
 FILING DATE: 19-SEP-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Vincent, Matthew P.  
 REGISTRATION NUMBER: 36,709  
 REFERENCE/DOCKET NUMBER: HMI-011CP2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 227-7400  
 TELEFAX: (617) 227-5941  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 200 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 PCT-US95-11869-4

Query Match 9.3%; Score 172; DB 5; Length 200;  
 Best Local Similarity 35.3%; Pred. No. 4,1e-07;  
 Matches 49; Conservative 16; Mismatches 58; Indels 16; Gaps 5;  
 QY 15 ALLLGLVGLVSGLSLEP-----VYVNSANKRFOAGSGYVLPQIGRLDLCPR 65  
 DB 7 AALLLAIVG-VCVMSDDPKVTSDRYAVYVNSNPRFH-RGDYTVESINDYLDYCPHY 64  
 QY 66 RPPGHSNNVEFYKLYLVGAQGRCEAPAPNLLTCDRPLD---DLRFTIKFOEYSP 122  
 DB 65 EEPFLP--AERMERVLYVMVNYEGHASCDBRQKGFKEWECNRPDSGPKFSEKQLFTF 122

QY 123 NMGHEFRSHDYIYIATS 141  
 DB 123 FSLGFEFRPGHEHYIYISAS 141

RESULT 43  
 US-08-240-124-4  
 Sequence 4, Application US/08240124  
 Patent No. 5516658  
 GENERAL INFORMATION:  
 APPLICANT: BECKMANN, M. P.  
 TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE  
 TITLE OF INVENTION: RECEPTOR HEK  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: IMMUNEX CORPORATION  
 STREET: 51 UNIVERSITY STREET  
 CITY: SEATTLE  
 STATE: WASHINGTON  
 COUNTRY: USA  
 ZIP: 98101  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: Apple Macintosh  
 OPERATING SYSTEM: Apple System 7.1  
 SOFTWARE: Microsoft Word for Apple, Version 5.1a  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/240,124  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/161,132  
 FILING DATE: 03-DEC-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/114,426  
 FILING DATE: 30-AUG-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/109,745  
 FILING DATE: 20-AUG-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: SEBSE, KATHRYN A.  
 REGISTRATION NUMBER: 32,172  
 REFERENCE/DOCKET NUMBER: 2814-C  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (206) 587-0430  
 TELEFAX: (206) 233-0644  
 TELEX: 756822  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 201 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-240-124-4

Query Match 9.2%; Score 169.5; DB 1; Length 201;  
 Best Local Similarity 29.9%; Pred. No. 6,7e-07;  
 Matches 66; Conservative 18; Mismatches 82; Indels 55; Gaps 10;

QY 24 LVSGLSL-EPVYVNSANKRFOAGSGYVLPQIGRLDLCPRAPPPGHSNNVEFYK 82  
 DB 20 LRGSSTLRHYVYVNSNPNRL-LRGDAVVELGNDYLDIVCPHYEGPPEGP--ETFA 76  
 QY 83 LVGAQGRCEAP-PAPNLLTCDRPLDRLRFTIKFOEYSPNLMGHEFRSHDYIYIATS 141  
 DB 77 MDMPGYESCGAGPAPYKRWVCSLPGHYQSEKIQRTTSPSLGTEFLRGTEYIYIS 136  
 QY 142 DGTREGLESIGGVCLTRGKYLRLRVGQSPRGAVRKPVESEPMERDRGAHSLPGE 201  
 DB 137 --TPE-----SSGQCL-----RLGVSVCKRKSSAPV----- 164

QY 202 NLPDPTSNATS--RGAEGPLPPSPMPAVAGAAGLALLL 240  
Db 165 ---GSPGSGTSGWRGCDTPSP-----LCILL 189

## RESULT 44

US-08-453-943-4  
Sequence 4, Application US/08453943  
Patent No. 5738844  
GENERAL INFORMATION:  
APPLICANT: BECKMANN, M. P.  
APPLICANT: CERRETTI, DOUGLAS P.  
TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE  
TITLE OF INVENTION: RECEPTOR HEX  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMUNEX CORPORATION  
STREET: 51 UNIVERSITY STREET  
CITY: SEATTLE  
STATE: WASHINGTON  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple System 7.1  
SOFTWARE: Microsoft Word for Apple, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,943  
FILING DATE: 30-MAY-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/240,124  
FILING DATE: 03-MAY-1994  
APPLICATION NUMBER: US 08/161,132  
FILING DATE: 03-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/114,426  
FILING DATE: 30-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/109,745  
FILING DATE: 20-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SEESE, KATHRYN A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2814-C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
TELEX: 756822  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 201 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-453-943-4

Query Match 9.2%; Score 169.5; DB 1; Length 201;  
Best Local Similarity 29.9%; Pred. No. 6.7e-07;  
Matches 66; Conservative 18; Mismatches 82; Indels 55; Gaps 10;  
QY 24 LVSGSL-EPVYNSANKRFOAEGGYLYPQIGRLDLCPRARPPGHSPPNVEFYKLY 82  
Db 20 LRGSSLRHYVYWNSSNPRL-LRGDAVVELGNDYLDIVCPHYEGGPPEG--ETFALY 76  
QY 83 LVGAAGRCEAP-PAPNLLTCDRPDLRFTIKFOEYSPULWGHERRSHDDYIATS 141  
Db 77 WMDWPGYSCQABGPAYRWKWCSPFGHVQSEKIQRTFSLGFEFLPGETIYIISVP 136  
QY 142 DGTREGLSLGGVCLTRGMKYLIRVGSPRGGAVERKPVSEMPERDRGAHSLPEPKE 201  
Db 137 --TPE-----SSGGL-----RLQVSVCKEKERSAHVP----- 164

QY 202 NLPDPTSNATS--RGAEGPLPPSPMPAVAGAAGLALLL 240  
Db 165 ---GSPGSGTSGWRGCDTPSP-----LCILL 189

## RESULT 45

US-09-057-121-4  
Sequence 4, Application US/09057121  
Patent No. 5969110  
GENERAL INFORMATION:  
APPLICANT: BECKMANN, M. P.  
APPLICANT: CERRETTI, DOUGLAS P.  
TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE  
TITLE OF INVENTION: RECEPTOR HEX  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMUNEX CORPORATION  
STREET: 51 UNIVERSITY STREET  
CITY: SEATTLE  
STATE: WASHINGTON  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple System 7.1  
SOFTWARE: Microsoft Word for Apple, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/057,121  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/240,124  
FILING DATE:  
APPLICATION NUMBER: US 08/161,132  
FILING DATE: 03-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/114,426  
FILING DATE: 30-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/109,745  
FILING DATE: 20-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SEESE, KATHRYN A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2814-C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
TELEX: 756822  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 201 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-057-121-4

Query Match 9.2%; Score 169.5; DB 2; Length 201;  
Best Local Similarity 29.9%; Pred. No. 6.7e-07;  
Matches 66; Conservative 18; Mismatches 82; Indels 55; Gaps 10;  
QY 24 LVSGSL-EPVYNSANKRFOAEGGYLYPQIGRLDLCPRARPPGHSPPNVEFYKLY 82  
Db 20 LRGSSLRHYVYWNSSNPRL-LRGDAVVELGNDYLDIVCPHYEGGPPEG--ETFALY 76  
QY 83 LVGAAGRCEAP-PAPNLLTCDRPDLRFTIKFOEYSPULWGHERRSHDDYIATS 141  
Db 77 WMDWPGYSCQABGPAYRWKWCSPFGHVQSEKIQRTFSLGFEFLPGETIYIISVP 136  
QY 142 DGTREGLSLGGVCLTRGMKYLIRVGSPRGGAVERKPVSEMPERDRGAHSLPEPKE 201  
Db 137 --TPE-----SSGGL-----RLQVSVCKEKERSAHVP----- 164

Db 137 -----SSGCL-----RLQVSVCKKRSASHV----- 164

QY 202 NLPDPTSNATS--RGAEGPLPPSPMPAVAGAAGLALLL 240

Db 165 ---GSPGESGTSGWRGDTSP-----LCULLL 189

RESULT 46

US-09-358-734-4

/ Sequence 4, Application US/09358734

/ Patent No. 6274117

/ GENERAL INFORMATION:

/ APPLICANT: BECKMANN, M. P.

/ TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE

/ TITLE OF INVENTION: RECEPTOR HER

/ NUMBER OF SEQUENCES: 4

/ CORRESPONDENCE ADDRESS:

/ ADDRESSEE: IMMUNEX CORPORATION

/ STREET: 51 UNIVERSITY STREET

/ CITY: SEATTLE

/ STATE: WASHINGTON

/ COUNTRY: USA

/ ZIP: 98101

/ COMPUTER READABLE FORM:

/ MEDIUM TYPE: Floppy disk

/ OPERATING SYSTEM: Apple System 7.1

/ SOFTWARE: Microsoft Word for Apple, Version 5.1a

/ CURRENT APPLICATION DATA:

/ APPLICATION NUMBER: US/09/358,734

/ FILING DATE:

/ CLASSIFICATION:

/ PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: 08/240,124

/ FILING DATE:

/ PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: US 08/114,426

/ FILING DATE: 30-AUG-1993

/ PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: US 08/109,745

/ FILING DATE: 20-AUG-1993

/ ATTORNEY/AGENT INFORMATION:

/ NAME: SEESE, KATHRYN A.

/ REGISTRATION NUMBER: 32,172

/ REFERENCE/DOCKET NUMBER: 2814-C

/ TELECOMMUNICATION INFORMATION:

/ TELEPHONE: (206) 587-0430

/ TELEFAX: (206) 233-0644

/ TELETYPE: 756822

/ INFORMATION FOR SEQ ID NO: 4:

/ SEQUENCE CHARACTERISTICS:

/ LENGTH: 201 amino acids

/ TYPE: amino acid

/ TOPOLOGY: linear

/ MOLECULE TYPE: protein

US-09-358-734-4

Query Match 9.2%; Score 169.5; DB 4; Length 201;

Best Local Similarity 29.9%; Pred. No. 6.7e-07; Indels 55; Gaps 10;

Matches 66; Conservative 18; Mismatches 82;

Db 24 LVSGLSL-EPYVNSANKRFQAEAGGYLYPQIGRLDLCPRAPPGPHSGPNYEFYKLY 82

Db 20 LRGSLSLHVYVYVNSNRL-LRQDAVELGLNDYLDIVCHYEGPPEGP--ETPALY 76

QY 83 LVGAGQGRCEAP-PAPLLLTCPRPDLRFTTKFQYSPNLMGHERSHDYIATS 141

Db 77 MVDWGVSCQAEGRAYKRWVCSLPFGHVPSERIQFTFPLSGFEPLPEETYYISVP 136

QY 142 DGTREGSLQGVCLTRGMKVLRLVQSPRGAVPRKPVSEMPERDRGAHSLPEQKE 201

Db 137 --TPE-----SSGCL-----RLQVSVCKKRSASHV----- 164

QY 202 NLPDPTSNATS--RGAEGPLPPSPMPAVAGAAGLALLL 240

Db 165 ---GSPGESGTSGWRGDTSP-----LCULLL 189

RESULT 47

US-09-214-631-8

/ Sequence 8, Application US/09214631

/ Patent No. 6413730

/ GENERAL INFORMATION:

/ APPLICANT: Holland, Sacha

/ APPLICANT: Mbamalu, Geraldine

/ APPLICANT: Pawson, Tony

/ TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED

/ TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR

/ FILE REFERENCE: 11757.23USMO

/ CURRENT APPLICATION NUMBER: US/09/214,631

/ CURRENT FILING DATE: 1999-03-12

/ EARLIER APPLICATION NUMBER: PCT/CA97/00473

/ EARLIER FILING DATE: 1997-07-04

/ EARLIER APPLICATION NUMBER: 60/021,272

/ EARLIER FILING DATE: 1996-07-05

/ NUMBER OF SEQ ID NOS: 13

/ SOFTWARE: Patent In Ver. 2.0

US-09-214-631-8

Query Match 9.1%; Score 169; DB 4; Length 201;

Best Local Similarity 29.8%; Pred. No. 7.4e-07; Indels 62; Gaps 11;

Matches 70; Conservative 19; Mismatches 84;

Db 17 LILGYL-----GLVSGLSL-EPYVNSANKRFQAEAGGYLYPQIGRLDLCPRAP 68

Db 6 LILTYVMAFLSPRGSSLSLHVYVNSNRL-LRQDAVELGLNDYLDIVCHYEGP 64

QY 69 GPHSPNVEFYLYLVGAQGRCEAP-PAPLLLTCPRPDLRFTTKFQYSPNLMG 127

Db 65 GPPEGP--ETPALYVMDWGVSCQAEGRAYKRWVCSLPFGHVPSERIQFTFPLSGF 122

QY 128 EFRSHDYIATS DGTREGSLQGVCLTRGMKVLRLVQSPRGAVPRKPVSEMPER 187

Db 123 EFLPEETYYISVP--TPE-----SSGCL-----RLQVSVCKK 155

QY 188 RDGAHSLPEQKENLPDPTSNATS--RGAEGPLPPSPMPAVAGAAGLALLL 240

Db 156 RKSESAPHV-----GSPGESGTSGWRGDTSP-----LCULLL 189

RESULT 48

US-09-214-631-6

/ Sequence 6, Application US/09214631

/ Patent No. 6413730

/ GENERAL INFORMATION:

/ APPLICANT: Holland, Sacha

/ APPLICANT: Mbamalu, Geraldine

/ APPLICANT: Pawson, Tony

/ TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED

/ TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR

/ FILE REFERENCE: 11757.23USMO

/ CURRENT APPLICATION NUMBER: US/09/214,631

/ CURRENT FILING DATE: 1999-03-12

/ EARLIER APPLICATION NUMBER: PCT/CA97/00473

/ EARLIER FILING DATE: 1997-07-04

/ EARLIER APPLICATION NUMBER: 60/021,272

/ EARLIER FILING DATE: 1996-07-05

/ NUMBER OF SEQ ID NOS: 13

/ SOFTWARE: Patent In Ver. 2.0





Wed Feb 12 09:30:14 2003

us-10-021-121-4.ra1

Page 23

Search completed: February 11, 2003, 12:06:26  
Job time : 14.6918 secs

---

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2003, 11:48:59 / Search time 24.805 Seconds  
(without alignment)  
1317.705 Million cell updates/sec

Title: US-10-021-121-4

Sequence: 1 MGPSPGPGVGVGALLLG.....PVIYQDGPQSPNIIYKV 340

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 50 summaries

Database: PIR73:\*

1: p1r1:\*

2: p1r2:\*

3: p1r3:\*

4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	637.5	34.5	336	2	149766
2	632	34.2	346	2	S46993
3	629.5	34.0	333	2	184743
4	613.5	33.2	345	2	148780
5	608.5	32.9	345	2	158406
6	214.5	11.6	237	2	119914
7	179	9.7	238	2	138849
8	176	9.5	209	2	A54984
9	175.5	9.5	213	2	JEO322
10	170.5	9.2	228	2	A57084
11	169.5	9.2	201	2	138850
12	167.5	9.1	228	2	158170
13	166	9.0	205	2	A36377
14	159	8.6	680	2	S31216
15	154.5	8.4	1049	1	CGH078
16	153.5	8.3	1670	1	CGH03B
17	151.5	8.2	301	2	B31129
18	149	8.1	325	2	T32248
19	149	8.1	569	2	S42886
20	148.5	8.0	316	2	T20497
21	148	8.0	921	2	S42617
22	146.5	7.9	1315	2	A56101
23	146.5	7.9	1452	2	A40333
24	146.5	7.9	1774	2	B56101
25	146	7.9	675	2	S20819
26	145.5	7.9	305	2	T20906
27	145	7.8	674	2	S13301
28	145	7.8	931	2	S13580
29	144.5	7.8	438	2	S53787

30	144	7.8	1027	2	S28774	collagen alpha cna
31	143	7.7	1747	2	A54121	collagen alpha-4 c
32	142.5	7.7	743	1	S23779	collagen alpha 1(V
33	142.5	7.7	1496	1	CGH02V	collagen alpha 2(V
34	142	7.7	744	2	S15435	collagen alpha 1(V
35	142	7.7	1029	1	S21369	collagen alpha 2(V
36	142	7.7	1763	2	S16366	collagen alpha 2(I
37	141.5	7.6	1466	1	CGH07L	collagen alpha 1(I
38	141	7.6	319	2	T32250	hypothetical prote
39	141	7.6	744	1	A34246	collagen alpha 1(V
40	141	7.6	744	1	S23298	collagen alpha 1(V
41	140.5	7.6	305	2	T30165	hypothetical prote
42	140	7.6	304	2	T22482	hypothetical prote
43	140	7.6	680	1	CGH01D	collagen alpha 1(X
44	139.5	7.5	210	2	B44984	collagen - nematod
45	139	7.5	940	2	JEO291	FB19 protein - hum
46	139	7.5	1024	2	S18251	collagen alpha 1(X
47	138.5	7.5	469	2	A24450	collagen alpha 2(V
48	138.5	7.5	1419	2	A41182	collagen alpha 1(I
49	138.5	7.5	1487	2	B41182	collagen alpha 1(I
50	138.5	7.5	1546	1	CGH02E	collagen alpha 2(X

#### ALIGNMENTS

RESULT 1  
149766  
hepatoma transmembrane kinase ligand - mouse  
C/Species: Mus musculus (house mouse)  
C/Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999  
C/Accession: 149766  
R/Bennett, B.D.; Zeigler, F.C.; Gu, Q.; Fendly, B.; Goddard, A.D.; Gillett, N.; Matthew  
Proc. Natl. Acad. Sci. U.S.A. 92, 1866-1870, 1995  
A/Title: Molecular cloning of a ligand for the EPH-related receptor protein-tyrosine ki  
A/Reference number: 149766; MUID:9519254; PMID:7534404  
A/Accession: 149766  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-336 <RES>  
A/Cross-references: GB:L38847; NID:g769677; PIDN:AAC42052.1; PID:g769678  
C/Genetics:  
A/Genes: HTK

Query Match 34.5%; Score 637.5; DB 2; Length 336;  
Best Local Similarity 42.1%; Pred. No. 1.9e-37;  
Matches 139; Conservative 49; Mismatches 129; Indels 13; Gaps 5;  
QY 14 GALLIGVGLVSGLSLEPVYVNSANKRFQAEGVVYPOIGDRLLDLCPRAPFGPHSS 73  
DB 17 GLMWLCRTAISRSIVLEPIYVNSNSKFLPGQGLVLPQIGDLIDICPKV---DSKIV 73  
QY 74 PNYEFYKLYVGAQGRRCAPAPNLLITCDPDLIRFTIKFOEYSPNIMGHEFRSH 133  
DB 74 GQYEVYKLYVMDKQADRCITKENTPILNCAIPDOVKFKIFQESFNMGLFQKNK 133  
QY 134 DYIYIATSDGTREGLESQGVGVCITRGKVLRYGO---SPRGAVPRKPVSEMER-DR 190  
DB 134 DYIYIISTNSGLEJDNQGGVCCOTRAKILMKYQDASSGASRNNGPTRRPELAGTN 193  
QY 191 GAHSLEFGKENLPDDPISNATSRGAEGPLPPSPMFAVGAAGGIALLLGVAGAGMC 250  
DB 194 GRSSITSPFVKNPSSSTIDNSAGHSNNLLGSEVALFAGIASCIIFIVIIITLVLL 253  
QY 251 WRRRAKSESRRHPGSGFGGSGIGLGGGGMGRPREPELDITARRGGAAPPPCGH 310  
DB 254 KYRRHRHRHSPOHTTTLSTLATPKGKNV---NGSEBPDVILPLK---TADSVFCFH 306  
QY 311 YEKVSGDYGHPIYIVQDGPQSPNIIYKV 340  
DB 307 YEKVSGDYGHPIYIVQDGPQSPNIIYKV 336

## RESULT 2

S46993  
 elik ligand - human  
 C/Species: Homo sapiens (man)  
 C/Date: 15-Jul-1995 #sequence\_revision 10-Nov-1995 #text\_change 28-May-1999  
 C/Accession: S46993  
 R/Beckmann, M.P.; Cerretti, D.P.; Baum, P.; Vanden Bos, T.; James, L.; Farrish, T.; Kozlowski, J. 13, 3757-3762, 1994  
 A/Title: Molecular characterization of a family of ligands for eph-related tyrosine kinase  
 A/Reference number: S46993; MUID:94349923; PMID:8070404  
 A/Accession: S46993  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-346 <BEG>  
 A/Cross-references: GB:U09304; NID:G538366; PIDN:AAA53093.1; PID:G538367

Query Match 34.2%; Score 632; DB 2; Length 346;  
 Best Local Similarity 39.5%; Pred. No. 4.7e-37;

Matches 145; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

```

QY 8 PGGVVGLLLGLVGLVSGL-----SLSPYVNSANKRFQAEAGVLYVQIGDRDLCLPRAPPPPHSS
DB 4 PQQRMLGKVLVAVVVALCRLATPLAKNLEFVSWSLNPKFLSGKGLVITPKIGDKDIT 63
QY 62 CPRAPPPPHSSPNYEFYLVVGAQGRRCRAPPNLLTCDRPDLLEFTIKFOEYS 121
DB 64 CPRAEAGR-----YEVYLVVREPEQAAGSTVLDPNVLTCKNPEGEIRFTIKFOEFS 118
QY 122 PNLNGHEFRSHDYITIAISDGTREGESLQGVCLTRGMKVLAVGQSPGQAVPRKPV 181
DB 119 PVMGJEFFKHHDYITTSNGSLGLENGEGVCRTKIMKIVGODPNNAVTPEDLTT 178
QY 182 SEMPMERDGAASLE-PEKENLPDPTSNATSRGAEGPLPPSPMAVAGAAGLA----- 236
DB 179 SRPSKADVTVMQAQAPSRGSLGSDSKHETVNGEKSGL-----GASGGSSGDPD 231
QY 237 -----LILLGVAGAGCA-----MCTRRRAKSESNSHPGSGRGSGSLGL 277
DB 232 GFENSKVALFAAVGAGCVFLIILFLVTLTLKLRKRHKHQQ-----RAAALSL 282
QY 278 -----GGGGGNGPREAPRGELGIALRGGAADPPCPHYEYSGDGHPIYIVQDPPQSP 333
DB 283 STASKSGSGTRAGTSPDITIIPLR---TTNNYCPHYEYSGDGHPIYIVQDPPQSP 339
QY 334 PNYYKYV 340
DB 340 ANIYYKV 346

```

## RESULT 3

I84743  
 hepatoma transmembrane kinase ligand - human  
 C/Species: Homo sapiens (man)  
 C/Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 05-Nov-1999  
 C/Accession: I84743  
 R/Bennett, B.D.; Zeigler, F.C.; Gu, Q.; Fendly, B.; Goddard, A.D.; Gillett, N.; Matthews  
 Proc. Natl. Acad. Sci. U.S.A. 92:1866-1870, 1995  
 A/Title: Molecular cloning of a ligand for the EPH-related receptor protein-tyrosine kin  
 A/Reference number: I49766; MUID:95199254; PMID:7534404  
 A/Accession: I84743  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-333 <RES>  
 A/Cross-references: GB:I38734; NID:G769675; PIDN:AA41752.1; PID:G769676  
 C/Genetic: GDB:EPUG5; LERK5  
 A/Gene: GDB:EPUG5; LERK5  
 A/Cross-references: GDB:438338; OMIM:600527  
 A/Map position: 13q33-13q33

Query Match 34.0%; Score 629.5; DB 2; Length 333;  
 Best Local Similarity 40.9%; Pred. No. 6.8e-37;  
 Matches 135; Conservative 52; Mismatches 130; Indels 13; Gaps 5;

QY 14 GALLILGLVGLVGLSEFVNSANKRFQAEAGVLYVQIGDRDLCLPRAPPPPHSS 73

DB 14 GVLAVLCRTAISKSYLSEFVNSANKRFQAEAGVLYVQIGDRDLCLPRAPPPPHSS 70

QY 74 PNVEFYKVLVGAQGRRCRAPPNLLTCDRPDLLEFTIKFOEYSNNGHEFRSH 133

DB 71 GQVEYKVVVVDQDQDRCTIKENTPILNCAKPPDDIFITIKFOEYSNNGHEFRSH 130

QY 134 DYTITNSGTREGESLQGVCLTRGMKVLAVGQSPGQAVPRKPV 180

DB 131 DYTITNSGTREGESLQGVCLTRGMKVLAVGQSPGQAVPRKPV 180

QY 191 GAASLEPEKKNLPDPTSNATSRGAEGPLPPSPMAVAGAAGLA----- 250

DB 191 GRSSTSPFKVPVPGSGTDSNAGSGNNILGSEVALFAGIASGCIIFITITLVLL 250

QY 251 WRRRRAPPSRRPFGSGSGSLGSGGGGMPREAPRGELTLRGGAADPPQSP 310

DB 251 KYRRRRKRSPOHTTTLSTATPRSGNN---NGSSPSDITIIPLR---TADSVCPH 303

QY 311 YEKVSGDGHPIYIVQDPPQSPPNYYKYV 340

DB 304 YEKVSGDGHPIYIVQDPPQSPPNYYKYV 333

RESULT 4

I48780  
 Stral/Ep1g2 protein - mouse

C/Species: Mus musculus (house mouse)  
 C/Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999

C/Accession: I48780; A55507; A55062; S52670

R/Boulic, P.; Ould-Abdelghani, M.; Vicaire, S.; Garnier, J.M.; Schunauer, B.; Dolle,

Dev. Biol. 170, 420-433, 1995

A/Title: Efficient cloning of cDNAs of retinoic acid-responsive genes in P19 embryonal

A/Reference number: I48780; MUID:9537533; PMID:7649373

A/Accession: I48780

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-345 <RES>

A/Cross-references: EMBL:248781; NID:G747858; PIDN:CAA8695.1; PID:G747859

R/Fletcher, F.A.; Renshaw, B.; Hollingsworth, T.; Baum, P.; Lyman, S.D.; Jenkins, N.A.,

Genomics 24, 127-132, 1994

A/Title: Genomic organization and chromosomal localization of mouse Ep1g2, a gene encod

A/Reference number: A55507; MUID:95203667; PMID:7896266

A/Accession: A55507

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-345 <FILE>

A/Cross-references: GB:U07598

R/Shao, H.; Lou, L.; Pandey, A.; Pasquale, E.B.; Dixit, V.M.

U. Biol. Chem. 269, 2606-2609, 1994

A/Title: cDNA cloning and characterization of a ligand for the Cdk5 receptor protein-ty

A/Reference number: A55062; MUID:95014510; PMID:7929389

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 1-89; T, 91-345 <SHA>

A/Cross-references: GB:U12883; NID:G575928; PIDN:AAA53231.1; PID:G575929

C/Genetic: GDB:EPUG2

Query Match 33.2%; Score 613.5; DB 2; Length 345;  
 Best Local Similarity 38.2%; Pred. No. 9.3e-36;  
 Matches 138; Conservative 51; Mismatches 107; Indels 65; Gaps 10;

QY 15 ALLLLGLVGLVSGL-----SLSPYVNSANKRFQAEAGVLYVQIGDRDLCLPRAPPPPHSS 72

DB 15 AVVVLTLCLATPLAKNLEFVSWSLNPKFLSGKGLVITPKIGDKDITCPRAEAGR-- 72

QY 73 SPNVEFYKVLVGAQGRRCRAPPNLLTCDRPDLLEFTIKFOEYSNNGHEFRSH 132

DB 73 ---YEVYKVLVREPEQAAGSTVLDPNVLTCKNPEGEIRFTIKFOEYSNNGHEFRSH 129

```

Oy 123 HDYIIATSDGREGLESLOGGVCLTRGKVLRYRGOSPRGGAIVRKRKVSSEMPMRDGA 192
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 130 HDYIITSNSGSLBEGLENREGVCRTRMKIYMKIGODP-NAYTEQLTSPSPRESNT 188
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 193 AHSLEPGKENTLPDPTSNATSRGAEP-----LPPSPMAVAGAAG-----LA 236
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 189 VKT-----ATQAPGRGSQSDSDGKHETVNOEKSFGPGAGGGSGSDSEFFNSK 236
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 237 LLLLVAGAGAG-----MCMRRRAKPSSESHRPPGSFGRGSLGL-----GG 279
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 237 VALPFAAVAGACITFLIIITFLTVLLIKLRKRKRKTQQ-----RAALSLSTLSP 287
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 280 GSGMGCREAPGELGIALRGGAADPFCPHYEKVSGDYGHVYIVDGPPOSPPNITYK 339
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 288 KGSSTAGTEPEDIIIPLR-----TTENNYCPHYEKVSGDYGHVYIVQEMFPOSPPNITYK 344
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 340 V 340
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 345 V 345
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

## RESULT 5

158406  
 LERK-2 - rat  
 C.Species: Rattus norvegicus (Norway rat)  
 C.Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 05-Nov-1999  
 C.Accession: 158406  
 R.Fletcher, F.A.; Carpenter, M.; Shilling, H.; Baum, P.; Ziegler, S.; Gimpel, S.; Hollnagel, E.;  
 Oncogene 9, 3241-3248, 1994  
 A.Title: LERK-2, a ligand for the receptor tyrosine kinase ELK, is evolutionarily conserved  
 A.Reference number: 158406; MUID:95022634; PMID:7936648  
 A.Accession: 158406  
 A.Status: preliminary; translated from GB/EMBL/DBJ  
 A.Molecule type: mRNA  
 A.Residues: 1-345 <RES>  
 A.Cross-references: EMBL:U07560; NID:G563118; PIND:AAA3092.1; PID:G563119  
 C.Genetics:  
 A.Gene: Eplg2

Query Match	32.9%	Score 608.5	DB 2	Length 345
Best Local Similarity	38.0%	Pred. No. 2.1e-35		
Matches 137	Conservative	52	Mismatches 107	Indels 65
				Gaps 10

[illegible]

## RESULT 6

T19914  
 Hypochemical protein C43F9.8 - *Caenorhabditis elegans*  
 C.Species: *Caenorhabditis elegans*  
 C.Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C.Accession: T19914  
 R.McIntire, B.  
 Submitted to the EMBL Data Library, November 1996  
 A.Reference number: Z19195  
 A.Accession: T19914  
 A.Status: preliminary; translated from GB/EMBL/DBJ  
 A.Molecule type: DNA  
 A.Residues: 1-237 <Wtl>  
 A.Cross-references: EMBL:D82262; PIDD:CA854195.1; GSPDB:GN00022; CESP:C43F9.8  
 C.Experimental source: clone C43F9  
 C.Genetics:  
 A.Gene: CESP:C43F9.8  
 A.Map position: 4  
 A.Introns: 32/2, 96/3, 214/1

Query Match	11.6%	Score 214.5;	DB 2;	Length 237;
Best Local Similarity	25.7%	Pred. No. 4.7e-08;		
Matches 53, Conservative	41;	Mismatches 83;	Indels 29;	Gaps 6;

```

QY 11 VRVGLLLLVGLVLS-GLSLEPYVMSANRFPQAE3-GYVLVYPOIGRLDLICPRAPP 68
Db 11 MQIAPIILLSTLFPFGIMARKIPDINIMISNDPIFDVSNTDHVISVHGVRVSRCPKSDOT 60
QY 69 GHSSPNVPEFYLYLVGACGRCRCAFPAPUILLTCRPPDLIRFTIKFOEYSPULMGHE 128
Db 61 G-----KYEISTIVVSDSEYDHCTL-SKPLVACGCDQOTINASINITYRSTRTPPGGFE 114
QY 129 FRSHEDYITIA-----TSDGTRSGLESLQGSGVCLTRGKMTLLRVGQ 169
Db 115 FQPGKMYETLISKEVDALIIYETANQIIPPGTSDGTLBGIIDRKGGLCTAKQMKIKFEVQ 174
QY 170 SPRGAVPRKPYSEMPEMDGAHS 195
Db 175 DRGRIENR--FAARTLKDRDAHS 198

```

## RESULT 7

138849  
 LERK-3 - human  
 C.Species: Homo sapiens (man)  
 C.Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 29-Sep-1999  
 C.Accession: 138849  
 R.Kozlovsky, C.J.; Maraskovsky, E.; McGrew, J.T.; Vandenberg, T.; Teepe, M.; Lyman  
 Oncogene 10, 299-306, 1995  
 A.Title: Islands for the receptor tyrosine kinases hek and elk: isolation of cDNA  
 A.Reference number: 138849; MUID:95140419; PMID:7838529  
 A.Accession: 138849  
 A.Status: preliminary; translated from GB/EMBL/DBJ  
 A.Molecule type: mRNA  
 A.Residues: 1238 <Res>  
 A.Cross-references: EMBL:U14187; NID:G642832; PIDD:AACS0078.1; PID:G642833  
 C.Genetics:  
 A.Gene: GDB:BPY3  
 A.Cross-references: GDB:438336; OMT:601381  
 A.Map position: 1q21-1q22  
 C.Superfamily: axon guidance signal protein

Query Match	9.7%	Score 179;	DB 2;	Length 238;
Best Local Similarity	28.4%;	Pred. No. 1.4e-05;		
Matches	65;	Conservative 24;	Mismatches 60;	Gaps 12;

  

QY	7	G	G	G	T	A	T	G	A	L	L	I	G	T	G	S	G	S	T	E	P	T	M	N	S	A	N	K	F	O	A	E	G	T	V	L	P	Q	I	G	D	R	L	I	C	P	--	R	64
DB	24	G	G	G	-----	G	A	L	G	-----	N	R	H	A	V	M	N	S	S	O	L	R	R	-----	G	T	T	V	A	N	V	A	N	D	I	D	I	C	P	H	A	T	N	67					
QY	65	A	R	P	G	H	S	S	P	-----	N	V	E	F	K	L	T	V	G	A	G	A	G	R	-----	C	E	A	P	A	N	L	I	L	T	C	R	P	D	-----	D	L	E	F	T	I	K	117	
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
DB	68	S	G	V	G	E	G	A	P	G	E	G	A	E	Q	V	L	T	V	M	S	R	N	G	R	T	N	A	G	G	K	-----	R	W	E	C	R	P	A	H	A	P	S	I	P	S	E	K	126

QY 118 QVSPNLMGHEFRSHHDYIIATSDGTREGLSLGQVCLTRGMKYLRLRVGSPRGAVP 177  
 DB 127 QRYAFSLIGYEFHAGHEYYIISTPTNLIH-----MKCLR--WKVFCVCASTSHSG--- 174  
 QY 178 RKPVSSEMP-----MERDRGAHSLR-----PKENILP 204  
 DB 175 EKPVPITLQFTMGPNVKINVLDEDEGNPQVPLKELKISGTSFPRKHLF 223

## RESULT 8

A54984  
 ERF-1 protein precursor - mouse  
 N:Alternate names: Cerk7 ligand  
 C:Species: Mus musculus (house mouse)  
 C>Date: 11-Nov-1994 #sequence\_revision 11-Nov-1994 #text\_change 29-Sep-1999  
 C:Accession: A54984; A53873  
 R:Cheng, H.J.; Flanagan, J.G.  
 Cell 79, 157-168, 1994  
 A:Title: Identification and cloning of ERF-1, a developmentally expressed ligand for the  
 A:Reference number: A54984; PMID:9500776; PMID:7522971  
 C:Accession: A54984  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-209 <CHP>  
 A:Cross-references: GB:U14941; NID:9558836; PIDN:AAA53636.1; PID:9558837  
 J:Shao, H.; Lou, L.; Pandey, A.; Verderame, M.F.; Stever, D.A.; Dixit, V.M.  
 J: Biol. Chem. 270, 3467-3470, 1995  
 A:Title: cDNA cloning and characterization of a Cerk7 receptor protein-tyrosine kinase 11  
 A:Reference number: A55873; PMID:95181289; PMID:7876076  
 C:Accession: A55873  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-209 <SHA>  
 A:Cross-references: GB:U14752; NID:9681886; PIDN:AAA68520.1; PID:9681887  
 C:Superfamily: axon guidance signal protein  
 C:Keywords: lipoprotein; membrane protein

Query Match 9.5%; Score 176; DB 2; Length 209;  
 Best Local Similarity 29.3%; Pred. No. 2e-05;  
 Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;

QY 33 VYVNSANKRFQAE-----GGVYLPIQIGDRDLDCPRARPPGPHSSPVEFYKLYVG 87  
 DB 35 VYVNSNRPFOVSAGVGGTYVEVSINDYLDIYCPHYGALP-PAEMERHILYVNGE 93  
 QY 88 QGRCEAPPAFLLITCDRPLD---DLRFTIKFQVSPNLMGHEFRSHHDYIIATSDGT 144  
 DB 94 GHASCDHQRQGRKWECKRPAAPGGLKFSKFLFTFSLGFEPFRPHEYYIISATP-- 151  
 QY 145 REGESLQGVCLTRGMKYLRLRVGSPRGAVPRKPVSEMPMERDRGAHSLFEGKENLP 204  
 DB 152 ---FNLYDRPCLR--LKVYV-----FTNETLY 174  
 QY 205 GDP-----TSNATSRGAG 218  
 DB 175 EAPPEIFTSNCSGLIG 192

## RESULT 9

J50322  
 ephrin-A2 - human  
 C:Species: Homo sapiens (man)  
 C>Date: 05-Feb-1999 #sequence\_revision 05-Feb-1999 #text\_change 21-Jul-2000  
 C:Accession: J50322  
 R:Adams, H.; Pedetour, F.; Grosgeorge, J.; Logtenberg, T.  
 Biochem. Biophys. Res. Commun. 252, 378-382, 1998  
 A:Title: Cloning, chromosomal mapping, and tissue expression of the gene encoding the huma  
 A:Reference number: J50322; PMID:99045414; PMID:9826538  
 C:Accession: J50322  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-213 <AAS>  
 A:Cross-references: GB:A0007292; NID:93688367; PIDN:CAA07435.1; PID:93688368

C:Superfamily: axon guidance signal protein

Query Match 9.5%; Score 175.5; DB 2; Length 213;  
 Best Local Similarity 36.8%; Pred. No. 2.2e-05;  
 Matches 43; Conservative 14; Mismatches 51; Indels 9; Gaps 3;

QY 33 VYVNSANKRFQAE-----EGVYLPIQIGDRDLDCPRARPPGPHSSPVEFYKLYVG 87  
 DB 39 VYVNSNRPFRHAGADDDGGTYVEVSINDYLDIYCPHYGALP-PAEMERHILYVNGE 97  
 QY 88 QGRCEAPPAFLLITCDRPLD---DLRFTIKFQVSPNLMGHEFRSHHDYIIATSDGT 141  
 DB 98 GHASCDHQRQGRKWECKRPAAPGGLKFSKFLFTFSLGFEPFRPHEYYIISATP 154

## RESULT 10

A57084  
 regulatory axon guidance signal protein RAGS precursor - chicken  
 C:Species: Gallus gallus (chicken)  
 C>Date: 03-Nov-1995 #sequence\_revision 03-Nov-1995 #text\_change 20-Jun-2000  
 C:Accession: A57084  
 R:Drescher, U.; Kremsner, C.; Handwerker, C.; Loeschinger, J.; Noda, M.; Bonhoeffer, F.  
 Cell 82, 359-370, 1995  
 A:Title: In vitro guidance of retinal ganglion cell axons by RAGS, a 25 kDa retinal prot  
 A:Reference number: A57084; PMID:95360980; PMID:7634326  
 C:Accession: A57084  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-228 <DRS>  
 A:Cross-references: GB:X90377; NID:91061113; PIDN:CAA62027.1; PID:9984118  
 C:Superfamily: axon guidance signal protein  
 C:Keywords: glycoprotein; membrane protein; phosphatidylinositol linkage  
 F:1-20/Domain: signal sequence #status predicted <SIG>

Query Match 9.2%; Score 170.5; DB 2; Length 228;  
 Best Local Similarity 28.9%; Pred. No. 5.3e-05;  
 Matches 73; Conservative 30; Mismatches 91; Indels 59; Gaps 13;

QY 16 LLLIGVLT-VGSLSEP-----VYVNSANKRFQAEAGVLYPIQIGDRDLDCR 64  
 DB 6 MLLAVPALMVCVRGQEPGRKAVADRYAVYVNSTPRFO-QGDYHIDVINCINYLIVFCP 64  
 QY 65 ARPPGHSSPVEFYKLYVG-----GAQGRCEAPPAFLLITCDRPLDRL 112  
 DB 65 YEDSVPEKXT--ERYLYVWTFDGISSCDHISGKRWCKRPHSPN-----GPK 113  
 QY 113 FTIKFQVSPNLMGHEFRSHHDYIIATSDGTREGLSLGQVCLTRGMKYLRLRVGQ 169  
 DB 114 FSEKFLFTFSLGFEPFRPHEYYIISATP-----CLK--LKVFR-- 159  
 QY 170 SPRGAVPRKPVSEMPMERDRGAHSLFEGKENLGDPTSNATSRGAEPLPPEKMP--A 227  
 DB 160 -PANSCKTIQHDVDFVNDVENSLEPADDTV--RESAPSRG-ENAAQTPIPLRL 214  
 QY 228 VAGAGLALLL 240  
 DB 215 LALTLFLAMLLI 227

## RESULT 11

I38850  
 LERK-4 - human  
 C:Species: Homo sapiens (man)  
 C>Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 29-Sep-1999  
 C:Accession: I38850  
 R:Kozlovsky, C.U.; Marakovsky, E.; McGrew, J.T.; Vandenbos, T.; Teepe, M.; Lyman, S.D.;  
 Oncogene 10, 299-306, 1995  
 A:Title: Ligands for the receptor tyrosine kinases hek and elk: isolation of cDNAs enc  
 A:Reference number: I38849; PMID:95140419; PMID:7838529  
 C:Accession: I38850  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-201 <RES>



A>Title: Characterization of the mouse type X collagen gene.  
 A/Reference number: S30127; MUID:93261348; PMID:8492743  
 A/Accession: S30127  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-12, 'F', '14-26', 'S', '28-247', 'L', '249-285', 'A', '287-305', 'F', '307-416', 'S', '418-499', 'I', 'R', 'Ape', 'S.S.', 'Selin', 'M.F.', 'Hayashi', 'M.', 'Olsen', 'B.R.', 'Eur. J. Biochem. 206, 217-224, 1992  
 A>Title: Cloning of the human and mouse type X collagen gene and mapping of the mouse  
 A/Reference number: 148299; MUID:92267014; PMID:1587271  
 A/Accession: 148299  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 52-247, 'L', '249-285', 'A', '287-305', 'F', '307-416', 'S', '418-499', 'L', '501-566', 'C', '568', 'A/Cross-references: EMBL:X65121; NID:950482; PIDN:CAA6237.1; PID:9667031  
 A/Summers, T.A.; Irwin, M.H.; Mayne, R.; Balian, G.  
 J. Biol. Chem. 263, 581-587, 1988  
 A>Title: Monoclonal antibodies to type X collagen. Biosynthetic studies using an antibody  
 A/Reference number: S26397; MUID:88087150; PMID:2826450  
 A/Accession: S26397  
 A/Molecule type: protein  
 A/Residues: 'SDGYFSQ', '24-26', 'KQ' <SUM>  
 A/Genetics:  
 A/Map position: 10  
 A/Introns: 51/3  
 A/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
 C/Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-680/Product: collagen alpha 1(X) chain #status predicted <MAT>  
 F:553-679/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 8.4%; Score 159; DB 2; Length 680;  
 Best Local Similarity 25.5%; Pred. No. 0.0011;  
 Matches 97; Conservative 30; Mismatches 108; Indels 146; Gaps 24;

QY 1 MGPPHSGGCV-RVCAALLLVGLVLSGLSPVYVMSANKRFAEGGYLYPQIGRLD 59  
 DB 211 IGPP--GPSGVGRGNGPFGQPGI-----KQGRFGKXG----- 244  
 QY 60 LCPAPAPGPHSSPNVYFYLYLVGAQGRCAAPAPNLLTCDDPDDLRTIKFGE 119  
 DB 245 ----PSGPGFGQGP-----GKQGR--EGIGFPAIGSPGQPGI----- 277  
 QY 120 YSPNLWGEFHSMDYIATSDGT---REGLSLOGGCVLTGKMYLLRVGSPRGGA 175  
 DB 278 --PGKXGHPGSPG-----IAGPPGAPGFGKQGLPLRG-----QRG----- 314  
 QY 176 VPRKRVSMPEMRDGAHSLPCKENLPGDPTSNATSRGAFGLPPSPMPAVAGAGGL 235  
 DB 315 LFGAPGA--KGERGPRG--HGPBPGLPSP--GNMGPGKGI-PGNHGLPGAKGEI 364  
 QY 236 ALLLVGAGAGAMCRRRRRAKP--SSSRHPG-----PGSGRGGSLGAGGGG 282  
 DB 365 G--LVGPAGPRGA--RGARPGGLDGKTYGPGLNGPKGNPGLPGQKDPVGGTPG 419  
 QY 283 M-----GPRAPRGLGIALRGGAADPPPCPHYKKSSEYGPVY 323  
 DB 420 LRGPVPGVANGVPGHNGEAPR-GEPIGPTR--GTPGPVGPVPPGSKGDPNP-- 472  
 QY 324 IVDDGP-----PQSP 334  
 DB 473 -GAPGPAGIATKGLNGPFGP 492

RESULT 15  
 CGBOTS  
 collagen alpha 1(III) chain - bovine  
 C/Species: Bos primigenius taurus (cattle)  
 C/Date: 04-Dec-1986 #sequence revision 04-Dec-1986 #text change 07-May-1999  
 C/Accession: A02862; A38001; A38002; A38003; A38004; A38005; S71946  
 R/Fietzek, P.P.; Allmann, H.; Rauterberg, J.; Henkel, W.; Wächter, E.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979

A>Title: The covalent structure of calf skin type III collagen. I. The amino acid sequ  
 A/Reference number: A02862; MUID:80026026; PMID:488906  
 A/Accession: A02862  
 A/Molecule type: protein  
 A/Residues: 1-242 <FIE>  
 R/Dewes, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979  
 A>Title: The covalent structure of calf skin type III collagen. II. The amino acid seq  
 A/Reference number: A38001; MUID:80026027; PMID:488907  
 A/Accession: A38001  
 A/Molecule type: protein  
 A/Residues: 243-422 <DEW1>  
 R/Bentz, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979  
 A>Title: The covalent structure of calf skin type III collagen. III. The amino acid seq  
 A/Reference number: A38002; MUID:80026028; PMID:488908  
 A/Accession: A38002  
 A/Molecule type: protein  
 A/Residues: 423-571 <BEN>  
 R/Lang, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979  
 A>Title: The covalent structure of calf skin type III collagen. IV. The amino acid seq  
 A/Reference number: A38003; MUID:80026029; PMID:488909  
 A/Accession: A38003  
 A/Molecule type: protein  
 A/Residues: 572-808 <LAN>  
 R/Dewes, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979  
 A>Title: The covalent structure of calf skin type III collagen. V. The amino acid sequ  
 A/Reference number: A38004; MUID:80026030; PMID:488910  
 A/Accession: A38004  
 A/Molecule type: protein  
 A/Residues: 809-947 <DEW2>  
 R/Allmann, H.; Fietzek, P.P.; Glanville, R.W.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 861-868, 1979  
 A>Title: The covalent structure of calf skin type III collagen. VI. The amino acid seq  
 A/Reference number: A38005; MUID:80026031; PMID:488911  
 A/Accession: A38005  
 A/Molecule type: protein  
 A/Residues: 948-1049 <ALT>  
 A/Experimental source: skin  
 R/Henkel, W.  
 Blochem. J. 318, 497-503, 1996  
 A>Title: Cross-link analysis of the C-telopeptide domain from type III collagen.  
 A/Reference number: S71946; MUID:95404897; PMID:8809038  
 A/Accession: S71946  
 A/Molecule type: protein  
 A/Residues: 87-106;107-1029;1037-1049 <HEN>  
 C/Comment: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are  
 C/Superfamily: The type III collagen molecule is a trimer of identical chains, linked to ea  
 C/Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyprol  
 F:1-1049/Product: collagen alpha 1(III) chain #status experimental <CAB>  
 F:1-14/Region: amino-terminal nonhelical telopeptide  
 F:15-1040/Region: helical  
 F:1587-589/Region: cell attachment (R-G-D) motif  
 F:752-754/Region: cell attachment (R-G-D) motif  
 F:875-877/Region: cell attachment (R-G-D) motif  
 F:878-880/Region: cell attachment (R-G-D) motif  
 F:935-937/Region: cell attachment (R-G-D) motif  
 F:1041-1049/Region: carboxyl-terminal nonhelical telopeptide  
 F:95-107;119;938;950/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F:107;950/Modified site: allylsite (Lys) #status predicted  
 F:107/Binding site: carbohydrate (Lys) (covalent) #status experimental  
 F:1040;1041/Disulfide bonds: interchain #status predicted

Query Match 8.4%; Score 154.5; DB 1; Length 1049;  
 Best Local Similarity 26.5%; Pred. No. 0.0036;  
 Matches 60; Conservative 9; Mismatches 74; Indels 83; Gaps 8;

QY 171 PPGGAYPRKPYVEMMRDGA-----ASHLEPGKEN 202  
 DB 688 PPGGSGPAPPGPGGVKGRGSGPGGAGAPGPGGPPGPGSGNGNPDPGSGSGAPGKG 747



QY 203 LPDPTSNAT-----SRGAGPLPPSPMVAAGAGLALLLVAGA 245  
 Db 748 PPGPPSNGAPSGPISGPKDSDGPPERKAPRPGPPGAPGLAGIAG-----LTGARGL 802  
 QY 246 GAGMCMERRRRAKP-----SBSRHPGP-----GSFRGSGSLGCGGGMGPREAPGE 292  
 Db 803 AGPPGMFGAPSGPPGQIKGNKPGPSGNGERGPDPGLPGLAGTAGEPRDGNPGS 862  
 QY 293 LGILRGG-----GADPPFCPEYKVSQDYGHPVYIVDGGP 330  
 Db 863 DGLFGRDGAAGAKDKRGKNSPPAP-----GAPGP-----GPP 896

## RESULT 16

CGH03B

collagen alpha 3(IV) chain precursor, long splice form - human  
 N/Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form  
 C/Species: Homo sapiens (man)  
 C/Date: 28-Oct-1994 #sequence, revision 03-Oct-1995 #ext\_change 22-Jun-1999  
 C/Accession: A54763; A43928; A44043; A45971; A39786  
 R/Mariyama, M.; Leinonen, A.; Mochizuki, T.; Tytgvaason, K.; Reiders, S.T.  
 J. Biol. Chem. 269, 23013-23017, 1994

A/Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression  
 A/Reference number: A54763; MUID:94364994; PMID:8083201  
 A/Accession: A54763  
 A/Molecule type: mRNA  
 A/Residues: 1-1670 <MAR>  
 A/Cross-references: GB:X80031; NID:9577563; PID:9577564  
 R/Experimental source: Kidney  
 R/Turner, N.; Mason, P.T.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.  
 J. Clin. Invest. 89, 592-601, 1992

A/Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the alpha  
 A/Reference number: A43928; MUID:92147878; PMID:1373499  
 A/Accession: A43928  
 A/Molecule type: mRNA  
 A/Residues: 1331-1524, 171526-1670 <TUR>  
 A/Cross-references: GB:M81379  
 R/Experimental source: Kidney  
 R/Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
 J. Biol. Chem. 267, 19780-19784, 1992

A/Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpasture  
 A/Reference number: A44043; MUID:93015826; PMID:1400291  
 A/Accession: A44043  
 A/Molecule type: DNA, mRNA  
 A/Residues: 1386-1670 <QUI>  
 A/Cross-references: GB:M92993; NID:9177895; PIDN:AA21610.1; PID:9177896  
 R/Note: sequence extracted from NCBI backbone (NCBI:115597)  
 R/Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
 J. Biol. Chem. 269, 17358, 1994

A/Reference number: A44738; MUID:94274734; PMID:8006044  
 A/Contents: annotation; extratum; correction to intronic sequence in A44043  
 R/Bernal, D.; Quinones, S.; Saus, J.  
 J. Biol. Chem. 268, 12090-12094, 1993

A/Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.  
 A/Reference number: A45971; MUID:93280184; PMID:8505332  
 A/Accession: A45971  
 A/Status: nucleic acid sequence not shown  
 A/Molecule type: mRNA  
 A/Residues: 1427-1444 <BER>  
 R/Note: sequence extracted from NCBI backbone (NCBI:133363); sequence incorrectly ident  
 R/Moriison, K.E.; Mariyama, M.; Yang-Feng, T.L.; Reiders, S.T.  
 Am. J. Hum. Genet. 49, 545-554, 1991

A/Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain of  
 A/Reference number: A39786; MUID:9135370; PMID:1882840  
 A/Accession: A39786  
 A/Molecule type: mRNA  
 A/Residues: 1453-1593, A', 1595-1670 <MOR>  
 A/Cross-references: GB:S55790; NID:9234418; PIDN:AA19637.1; PID:9234419  
 R/Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
 ed and subsequently O-glycosylated.  
 C/Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope

## C/Genetics:

A/Genes: COL4A3  
 A/Cross-references: GDB:128351; OMIM:120070

A/Map position: 2q36-2q37  
 A/Introns: 1385/1; 1418/1; 1486/1; 1547/2; 1585/3; 1643/2 #status incomplete

A/Note: The alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands wit  
 C/Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3  
 mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric  
 er associations in the interrupted helical domain (with disulfide and desmosine cross-1

C/Function:  
 A/Description: minor structural component of extracellular basement membrane in kidney  
 C/Superfamily: collagen alpha 1(IV) chain  
 C/Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extrace

F/1-28/Domain: signal sequence #status predicted <SIG>  
 F/29-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MAT>  
 F/29-42/Domain: amino-terminal nonhelical, NH1 <NH1>  
 F/43-1438/Region: interrupted helical  
 F/791-793/Region: cell attachment (R-G-D) motif  
 F/996-998/Region: cell attachment (R-G-D) motif  
 F/1154-1156/Region: cell attachment (R-G-D) motif  
 F/1306-1308/Region: cell attachment (R-G-D) motif  
 F/1345-1347/Region: cell attachment (R-G-D) motif  
 F/1432-1434/Region: cell attachment (R-G-D) motif  
 F/1439-1670/Domain: carboxyl-terminal nonhelical, NCI <NCI>  
 F/1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>  
 F/1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>  
 F/3133-3339/41,125,422,476,479,682,722,809,1387/Disulfide bonds: interchain #status pred  
 F/253/Binding site: carboxylate (Asn) (covalent) #status predicted  
 F/1460-1548,1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted  
 F/1505-1511,1516-1622/Disulfide bonds: #status predicted  
 F/1570-1662,1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query/Match 8.3%; Score 153.5; DB 1; Length 1670;  
 Best Local Similarity 27.0%; Pred. No. 0.007;  
 Matches 108; Conservative 26; Mismatches 155; Indels 111; Gaps 25;

QY 2 GPPHSGPG-----VRVALLLGLVGLVSLSPVYVMSANR-----FOAEGGYLYP 52  
 Db 991 GPP--GPRDGLASTGNPBPGLLGPSSMGKMP-----GSKRGKILGFRGARGRPLP 1044  
 QY 53 QI-----GDRLLDLP-----RARPGRHSSPNTEFYKLVYVGAQGRCE----- 93  
 Db 1045 GHHGQDQKGE-----PGYSEGRPPGPPTGDPG-----LPDMGKGMGQGPFGH 1093

QY 94 -----APRPNL--LLTCDRPLDLRFT--IKQESVPLMNGE-----FRSHHDYII 138  
 Db 1094 LPPAPPEAGAPSGSPGPGPGKPPGPDGFGKIGKLLGPPGIRGPPGLPGSPBPBGKI 1153

QY 139 ATSDTRGLLESLQGVCLTRGMKVLTRVQSPRG--GAVPRKVSSEMPERDRGAHSL 196  
 Db 1154 RQDQG-RDGIPEPAG-----EKGETGLLAPPGPRGNPQAQAK-----GDRGA----- 1196

QY 197 EFGKENLRQDPTSNATNSGARGPLPPSPMVAAGAGLALLLVAGAAGMCMERRRA 256  
 Db 1197 -GFPGLPG-----RKAAGDAGPRGPTIRGPPGPGIAGIIPQGTNRPP--PGSRG 1248

QY 257 KPSERHHPG-----GSFRGSGSLG-LQGGGMGP--REARPGELGLRGG 301  
 Db 1249 SFGARPPPPPSHVITGKDKGSGHGRPKGPPPTAGDMPPRLGAGRTPLG--GP 1305

QY 302 ADPPFCPEYKVSQDYGHPVYIVDGP-----PQSPNII 336  
 Db 1306 RQDPGF--GQFPVKGKXNPGFLGSGIPGCPRIQPGPPGV 1344

RESULT 17  
 B31219  
 collagen 2 - Caenorhabditis elegans  
 C/Species: Caenorhabditis elegans  
 C/Date: 30-Jun-1989 #sequence, revision 30-Jun-1989 #ext\_change 21-Jul-2000  
 C/Accession: B31219; T37289; T26033; T37288  
 R/Kramer, J.M.; Cox, G.N.; Hiren, D.  
 Cell 30, 599-606, 1982

A>Title: Comparisons of the complete sequences of two collagen genes from *Caenorhabditis*  
 A>Reference number: A30826; MUID:8350544; PMID:7139711  
 A>Accession: B31219  
 A>Molecule type: DNA  
 A>Residues: 1-301 <KRA>  
 A>Cross-references: GB:V00148; NID:g6683; PIDN:CAA23464.1; PID:g6684  
 A>Accession: T37289  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A>Molecule type: DNA  
 A>Residues: 1-301 <KR>  
 A>Cross-references: EMBL:V00148; NID:g6683; PIDN:CAA23464.1; PID:g6684  
 R>Coles, L.  
 submitted to the EMBL Data Library, December 1995  
 A>Reference number: Z20141  
 A>Accession: T26033  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A>Molecule type: DNA  
 A>Residues: 1-301 <WIL>  
 A>Cross-references: EMBL:Z68301; PIDN:CAA92620.1; GSPDB:GN00022; CESP:W01B6.7  
 A>Experimental source: clone W01B6  
 R>Kramer, J.M.; Cox, G.N.; Hirsch, D.  
 J. Biol. Chem. 260, 1945-1951, 1985  
 A>Title: Expression of the *Caenorhabditis elegans* collagen genes col-1 and col-2 is develo  
 A>Reference number: Z21668; MUID:85105075; PMID:2578467  
 A>Accession: T37288  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A>Molecule type: DNA  
 A>Residues: 1-301 <KR2>  
 A>Cross-references: EMBL:J01048; NID:g156261; PIDN:AAA7990.1; PID:g156262  
 C>Genetics:  
 A>Gene: col-2  
 A>Map position: 4  
 A>Introns: 94/1  
 C>Superfamily: unassigned collagens

Query Match: 8 2%; Score 151.5; DB 2; Length 301;  
 Best Local Similarity 28.7%; Pred. No. 0.0015;  
 Matches 50; Conservative 15; Mismatches 72; Indels 37; Gaps 7;

QY 168 GQSPRGAVPRPVSEMPERDRG-----AASLSEPKENLPGDPT--SNATSRGAE 217  
 Db 129 GNPGRGASAPCPVTPGPPCPGPGPPGAPGPPGPPGCGNSGSPAGSGPPAGP 188  
 QY 218 GPUPPSPNPAVGAAGGALLLLGVAG--CGAMCWRRRRAKPSGSNHPG-GSFGRCG 273  
 Db 189 GPAGPAGNDGAGAGAGCG-----PGEPGASEQSGP-----GEPPAPGPPGAPGAGNDG 236  
 QY 274 SGLGGGGGGMGR-----EAPGELGIALRGGAADPPCPHYEKVSG 316  
 Db 237 AGCTGGPGAGGKGPAGGAGAGGADGPNPGPGTAKPGGEGKGLCPYCAIDG 290

RESULT 18  
 T32248  
 A>Hypothetical protein T15B7.4 - *Caenorhabditis elegans*  
 C>Species: *Caenorhabditis elegans*  
 C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000  
 C>Accession: T32248  
 R>Pauley, A.; Gattung, S.  
 submitted to the EMBL Data Library, September 1997  
 A>Description: The sequence of *C. elegans* cosmid T15B7.  
 A>Reference number: Z21139  
 A>Accession: T32248  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A>Molecule type: DNA  
 A>Residues: 1-315 <PAU>  
 A>Cross-references: EMBL:AF022985; PIDN:AA869961.1; GSPDB:GN00023; CESP:T15B7.4  
 A>Experimental source: strain Bristol N2; clone T15B7  
 C>Genetics:  
 A>Gene: CESP:T15B7.4  
 A>Map position: 5  
 A>Introns: 266/1  
 C>Superfamily: unassigned collagens

Query Match 149; 8.1%; Score 149; DB 2; Length 325;  
Best Local Similarity 26.0%; Pred. No. 0.0025;  
Matches 92; Conservative 25; Mismatches 109; Indels 128; Gaps 23;

13 VQALLILGLVGLVGLSTL-----EPVYNSANKRFQAEQ---YVLVQIGDRIDLILC 62  
1 MSASTLTVTASAAAGIAIVCFVTQGMIFNDINSFDEKIGELKEFGYEOIA--WQAMI 58  
63 PPARPPGPHSPNVEFYKLYWG-----GAQGRCEA-PPAPNLLTCDRPDLIL 111  
59 PTTRESSGSS-----FLGRKKRQACNCGAOSRCCPAPFP-----96  
112 RFTIKFQEYSPNLWGHFEFSHHDIYIATSDGTREGLESLOG--GYCLTRGMKVLIRVGO 169  
97 ----PGQPAPAPQEQHPG-----LAQQPSGARINPATGRPGFCIT-----123  
170 SPRGAVPRKPVSEMPMERDRGAHSLPEKENTLPDPTSNATSRGAEGPLPPSPMAVA 229  
134 CPAPAPPGPAPG-----PGA-----PPKGNNGPAPAPQS--GGRP-PGPRP--A 175  
230 GAAGGIALLLGLVAGAGAMCWRARRAKPSESHPG-PGSFGRG--SLGLGGGM---283  
176 GDAG-----SPGQGHFPGSPFNPRGGQRSGLPSPSRPGR 212  
284 -GPRAPPEHLGIALRGGAAP-PCPHYEKVSQDGHPIYIVQDPPQSPN 335  
213 PGP-AGPGQPQ---HSGAGSPGPGP---PGSGPGHSGNDGVGAPFN 257

RESULT 19  
542886  
collagen - silkworm  
CISpecies: Bombyx mori (silkworm)  
CDate: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 20-Sep-1999  
CAccession: 542886  
RCharEyre, P.P.; Beeson, M.M.; Fourche, J.J.; Bosquet, G.G.  
submitted to the EMBL Data Library, March 1994  
ADescription: A novel collagen encoding mRNA is transiently accumulated during ecdysis  
AReference number: 542886  
AAccession: 542886  
Astatus: preliminary  
AMolecule type: mRNA  
AResidues: 1-569 <CHA>  
ACross-references: EMBL:Z30348; NID:G457768; PIDN:CAA83002.1; PID:G457769  
Csuperfamily: unassigned collagens

Query Match 8.1%; Score 149; DB 2; Length 569;  
Best Local Similarity 29.3%; Pred. No. 0.0045;  
Matches 51; Conservative 14; Mismatches 77; Indels 32; Gaps 6;

171 PRGAVPRKPVSEMPMERDRGAHSLPEKENTLPDPTSNATSRGAEGPLPPSPMAVAG 230  
1 PGQGGQPIKPG-----PGYFGPGQPGYVPGP--GQDAPGPGPGQPGQGPTRPG 50  
231 MAGGALLLLGVAAGGAMCWRARRAKPSESHPG-----PGSRGRGSJLGGGGM 283  
51 QAG-----QPGYFGGG-----QPIKAPQHPGQPGQPGPGPGLPGQPGYFPQPGQ 99  
284 GREAPPEGLGIALRG-----GGAADPPPCPHYEKVSQDGHPIYIVQDPPQSP 333  
100 PGQPGQPGQGGYFGQGGPGQGGQPINPSPGVFGSGGQGYVGGQVQPGPQPGP 153

RESULT 20  
T00497  
hypochemical protein F02D10.1 - Caenorhabditis elegans  
CISpecies: Caenorhabditis elegans  
CDate: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
CAccession: T00497  
R:Swindburne, J  
submitted to the EMBL Data Library, November 1995  
AReference number: Z19283

A:Accession: T20497  
 A:Status: preliminary; translated from GB/EMBL/DBS  
 A:Molecule type: DNA  
 A:Residues: 1-316 <WIL>  
 A:Cross-references: EMBL:Z67990; PIDN:CAA91932.1; GSPDB:GNO0028; CESP:F02D10.1  
 A:Experimental source: clone F02D10  
 C:Genetics:  
 A:Gene: CESP:F02D10.1  
 A:Map position: X  
 A:Introns: 56/3  
 C:Superfamily: unassigned collagens

Query Match 8.0%; Score 148.5; DB 2; Length 316;  
 Best Local Similarity 28.5%; Pred. No. 0.0026;

Matches 49; Conservative 12; Mismatches 78; Indels 33; Gaps 6;

QY 168 GGGPRGAGVPRKVSMPERDRGAHSLPCKENLPDP-----TSNATSGA 216  
 DB 144 GASGKGAAPCEPSTPPQ-PCPAGPPPPGDPGPBPAGSPAGSPAGSPAGP 202  
 QY 217 EGPLPPSPPAVAGAGLALLLVAGAGAMCRRRAKSESRRPQP-GSPRGSL 275  
 DB 203 PGAGAPGNDGQPGQPGG--QDGAASAG-----EAGCPAGPPGPPAGPAGP 253  
 QY 276 GUGGGGCGMPR-----EAPFELGIALRGGAADPPCTHYEKVSG 316  
 DB 254 GSGSAGGPPKPPGPPAGQPGSDGNPTAGPPGNGEGEKICPKYCAIDG 305

## RESULT 21

842617  
 collagen alpha 1(IX) chain - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 25-Dec-1994 #sequence\_revision 19-Apr-1996 #text\_change 20-Sep-1999  
 C:Accession: S42617

R:Rokos, I.; Muragaki, Y.; Warman, M.; Olsen, B.R.

Matrix Biol. 14, 1-8, 1994  
 A:Title: Assembly and sequencing of a cDNA covering the entire mouse alpha-1(IX) collagen  
 A:Reference number: S42617; WUID:94340193; PMID:8061915  
 A:Accession: S42617  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-921 <ROK>  
 A:Cross-references: GB:L12215; GB:L19396; NID:9192675; PIDN:AAA21834.1; PID:9192676  
 C:Superfamily: unassigned collagens

Query Match 8.0%; Score 148; DB 2; Length 921;  
 Best Local Similarity 24.0%; Pred. No. 0.009;  
 Matches 87; Conservative 23; Mismatches 130; Indels 122; Gaps 17;

QY 2 GPPHSGPGVRYGALLLVGLVSLSPYVNSANKRFQABGGYLYPQIGDRDLL 61  
 DB 391 GPP-GRPPS-----GTIGHDQ-----DPL 410  
 QY 62 CPARPSPGSHSPNYFYKLYLVG-----GAQRCEAPPAENLLTCRPLDLRTFK 116  
 DB 411 CPNSCPGSSGYG-----LPKMGHGKAGKEIG----- 440  
 QY 117 FQRYSNLWGEFRSHHDYITATSDTREGLESLOG--GVCLTGMKYLVRVSGSPRG 174  
 DB 441 -----FGRQGHKEEDQDELGEVDQCPGPGQGLRGITGIVGDGEKARFDEPQPQ 495  
 QY 175 AVP-----RKVSSEMERDRGAHSLPCKENLPDPDSNMTS--RGAEGPLPPPS 224  
 DB 496 GITGAAGDQGRCPGEPETGPEBDRGT-----QSRCTPSPSPKPKDPTGLPGVDGDIPLG 550  
 QY 225 MPVAVAG-----GALL--LVGAGAGAMCWRARRAKPSE-----SRHPG--- 265  
 DB 551 MPSTKEAGKPPGPDVGIAGLPVPGIPGAKGVAGEKNTGAPGPGQLGSSGKPGQGG 610  
 QY 266 -PSPFRGSLG--GGGGGAGPREAE--RGEIGIALRG--GAADPPFPHYKXVSGYX 319  
 DB 611 PPEVPPRGRLGPGSRGVPVGPBGSPGIPGKUGSVGSPGLPLGPPPLGKMGKGRDVG 670

QY 320 HP 321  
 DB 671 EP 672

## RESULT 22

A56101  
 collagen alpha 1(XVIII) chain precursor, short splice form - mouse

N:Contains: endostatin  
 C:Species: Mus musculus (house mouse)

C>Date: 03-Oct-1995 #sequence\_revision 08-May-1998 #text\_change 31-Mar-2000  
 C:Accession: A56101; A58371; S72450; S65595

R:Rehm, M.; Philjaient, T.  
 J. Biol. Chem. 270, 4705-4711, 1995

A:Title: Identification of three N-terminal ends of type XVIII collagen chains and tissue homologous to rat and Drosophila fibrinogen proteins.  
 A:Reference number: A56101; WUID:95181468; PMID:7876242

A:Accession: A56101  
 A:Molecule type: mRNA

A:Residues: 1-103 <REH1>  
 A:Cross-references: GB:L11636; NID:618427; PIDN:AAC52178.1; PID:618428

R:Rehm, M.; Philjaient, T.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994

A:Title: Alpha1(XVIII), a collagen chain with frequent interruptions in the collagenous  
 A:Reference number: A58371; WUID:94240112; PMID:8183994

A:Accession: A58371  
 A:Molecule type: mRNA

A:Residues: 1-928 <REH2>  
 A:Cross-references: GB:L11698; NID:9404754; PIDN:AAA7434.1; PID:9553894

R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.  
 submitted to the EMBL Data Library, August 1993

A:Reference number: S72450  
 A:Accession: S72450

A:Molecule type: mRNA  
 A:Residues: 26-687, 'L', 689-734, 'P', 736-751, 'R', 753-1315 <OHS>

A:Cross-references: EMBL:422545; NID:9348968; PIDN:AAA19787.1; PID:9511298  
 R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.

Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994

A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xa  
 A:Reference number: A58370; WUID:94240111; PMID:8183993

A:Accession: S65595  
 A:Molecule type: mRNA

A:Residues: 26-1315 <OHS>  
 A:Cross-references: EMBL:422545

C:Comment: Prolines and Iyenes at the third position of the tripeptide repeating unit  
 lated and subsequently O-glycosylated.

C:Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in per  
 C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of u

ay be useful in treating solid tumors.  
 C:Genetics:

A:Gene: MGI:Coll8a1  
 A:Cross-references: MGI:71175

A:Map position: 10:41.0  
 C:Superfamily: unassigned collagens

C:Keyworder: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc

F:1-25/Domain: signal sequence #status predicted <SIG>  
 F:24-225/Region: thrombospondin amino-terminal similarity

F:26-1315/Product: collagen alpha 1(XVIII) chain, short splice form #status predicted <  
 F:337-153/Domain: collagenous #status predicted <CO1>

F:342-437/Domain: collagenous #status predicted <CO2>  
 F:462-583/Domain: collagenous #status predicted <CO3>

F:607-689/Domain: collagenous #status predicted <CO4>  
 F:704-745/Domain: collagenous #status predicted <CO5>

F:759-831/Domain: collagenous #status predicted <CO6>  
 F:842-874/Domain: collagenous #status predicted <CO7>

F:887-910/Domain: collagenous #status predicted <CO8>  
 F:892-894/Region: cell attachment (R-G-D) motif

F:918-969/Domain: collagenous #status predicted <CO9>  
 F:993-1000/Domain: collagenous #status predicted <CO10>

F:1132-1315/Product: endostatin #status predicted <EST>  
 F:1139-1315/Region: multiplexin collagen carboxyl-terminal similarity

F:116-488/Binding site: carbohydrate (Asn) (covalent) #status predicted



A:Map position: 10:41.0  
 A:Intons: 1295/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 1599/3  
 A>Note: the 1st of introns is incomplete  
 C:Superfamily: unassigned collagens  
 C:Keyword: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc  
 F:1-1774/Domain: collagen alpha 1(XVII) chain precursor, long splice form #status pred  
 F:1-239/487-1774/Domain: collagen alpha 1(XVII) chain precursor, medium splice form #s  
 F:1-24/Domain: signal sequence #status predicted <Sig>  
 F:361-486/Region: frizzled similarity  
 F:786-812/Domain: collagenous #status predicted <CO01>  
 F:823-896/Domain: collagenous #status predicted <CO02>  
 F:921-1042/Domain: collagenous #status predicted <CO03>  
 F:1066-1148/Domain: collagenous #status predicted <CO04>  
 F:1163-1204/Domain: collagenous #status predicted <CO05>  
 F:1218-1290/Domain: collagenous #status predicted <CO06>  
 F:1301-1333/Domain: collagenous #status predicted <CO07>  
 F:1346-1369/Domain: collagenous #status predicted <CO08>  
 F:1351-1383/Region: cell attachment (R-G-D) motif  
 F:1377-1448/Domain: collagenous #status predicted <CO09>  
 F:1442-1459/Domain: collagenous #status predicted <CO10>  
 F:1591-1774/Domain: endostatin #status predicted <EST>  
 F:1598-1774/Region: multiplexin collagen carboxyl-terminal similarity  
 F:354-361-947/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:699-704-1716/Binding site: carbohydrate (Ser) (covalent) #status predicted  
 F:910-913-1053/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 7.9%; Score 146.5; DB 2; Length 1774;  
 Best Local Similarity 23.7%; Pred. No. 0.023;  
 Matches 92; Conservative 22; Mismatches 107; Indels 167; Gaps 21;

QY 2 GPHSGPGGVAVALLLGLG-LVSGLSLEPYNNNSANKRFAEGGYLYPQIGRDL 60  
 DB 798 GPP-GRAPQ-----GPAGVVOGSPNSQV----- 821

QY 61 LCPRAR-PPGPHSSPNVEFYKLYVGAQGRCE----- 93  
 DB 822 --PQAQPPGPQGPQPGKD-----GTPGRDGPDPGDPGRPGDTGPQGPFGTRGDVG 871

QY 94 -----APPANLLTCDPDDLDRTTTFQESPLVMKHERSHHDYITAT 140  
 DB 872 PKGEGDPGIGRPPGP-----PQPPSPSRQDXTFFIDWE 908

QY 141 SDGTREGLESLQGVCLTRGMKYLVRVQSPRGAVPRKP--VSEMPMERDR-GAHSL 197  
 DB 909 GSGFSGGIESLRG-----PRGFPPGPFGVPLGPGRFGINGSTA 951

QY 198 PKXENLPDPTSNATSRGAEP--LP-PPMPAVVAGAAGLALLLGVAGAGCMWRER 254  
 DB 952 PGAGLPGV-----GKEGPPGFPGPFPGPKEGP-----PGVAGQKGSV----- 994

QY 255 RAMPSSRRHGP-GSPRGSGSLGCGGGMGPRAEPGELIALRG--GAADPPCPHY 311  
 DB 995 ----GDVGLPDPGSKSDLPIDMPKSGLA---GSPGVPVGPFGPPGPPGPAAG 1047

QY 312 EKVSQDYGHFVIVQ-----DGPQSP 333  
 DB 1048 DMWEGS-GIPLMTTARSSDGLQGPSP 1074

RESULT 25  
 S20819  
 collagen alpha 3(IX) chain precursor - chicken  
 C:Species: Gallus gallus (chicken)  
 C:Date: 22-Nov-1993 #sequence revision 09-Mar-1996 #text change 20-Sep-1999  
 C:Accession: S20819; S22429; S22918; S22238; C18856; S22241  
 R:Breton, R.G.; Ouepenkaka, M.V.; van der Reest, M.; Wayne, R.  
 Eur. J. Biochem. 205, 443-449, 1992  
 A>Title: Cloning of the chicken alpha-3(IX) collagen chain completes the primary structu  
 A:Reference number: S20819; MUID:92241276; PMID:1572350  
 A:Accession: S20819  
 A:Molecule type: mRNA  
 A:Residues: 1-675 <ERR>  
 A:Cross-references: EMBL:X64712; NID:g63316; PIND:CAA4567.1; PID:g63317

R:Har-El, R.; Sharma, Y.D.; Aguilera, A.; Veyama, N.; Wu, J.J.; Eyre, D.R.; Juricic, L.  
 J. Biol. Chem. 267, 10070-10076, 1992  
 A>Title: Cloning and developmental expression of the alpha3 chain of chicken type IX co  
 A:Reference number: S22429; MUID:92230586; PMID:1577778  
 A:Accession: S22429  
 A:Molecule type: mRNA  
 A:Residues: 1-195, 'G', 197-675 <HA1>  
 A:Cross-references: EMBL:M83179  
 A>Note: 353-Arg, 386-Leu and 548-Arg were also found  
 R:Har-El, R.; Sharma, Y.D.; Aguilera, A.; Ueyama, N.; Wu, J.J.; Eyre, D.R.; Juricic, L.  
 submitted to the EMBL Data Library, February 1992  
 A:Description: Cloning and developmental expression of the alpha 3 chain of chicken typ  
 A:Reference number: S22918  
 A:Accession: S22918  
 A:Molecule type: mRNA  
 A:Residues: 1-195, 'G', 197-405, 'S', 407-675 <HA2>  
 A:Cross-references: EMBL:M83179; NID:g211040; PIND:AA55960.1; PID:g211041  
 R:Wayne, R.; van der Reest, M.; Nimomaya, Y.; Olsen, B.R.  
 Ann. N. Y. Acad. Sci. 460, 38-46, 1985  
 A>Title: The structure of type IX collagen.  
 A:Reference number: S22238; MUID:86185164; PMID:3868958  
 A:Accession: S22238  
 A:Molecule type: protein  
 A:Residues: 540-548 <MA>  
 R:Nimomaya, Y.; van der Reest, M.; Wayne, R.; Lozano, G.; Olsen, B.R.  
 Biochemistry 24, 4223-4229, 1985  
 A>Title: Construction and characterization of cDNA encoding the alpha2 chain of chicken  
 A:Reference number: A18856; MUID:86026268; PMID:2996593  
 A:Accession: C18856  
 A:Molecule type: protein  
 A:Residues: 540-558 <NI>  
 R:Shimomaki, M.; Wright, D.W.; Irwin, M.H.; van der Reest, M.; Wayne, R.  
 Ann. N. Y. Acad. Sci. 580, 1-7, 1990  
 A>Title: The structure and macromolecular organization of type IX collagen in cartilage  
 A:Reference number: S22241; MUID:90247791; PMID:2186687  
 A:Accession: S22241  
 A:Molecule type: protein  
 A:Residues: 135, 'E', 137-187, 'X', 189-191 <SR>  
 C:Superfamily: unassigned collagens  
 C:Keywords: coiled coil; connective tissue; disulfide bond; extracellular matrix; heter  
 F:1-21/Domain: signal sequence #status predicted <Sig>  
 F:22-675/Domain: collagen alpha 3(IX) chain #status predicted <MA>  
 F:22-24/Domain: non-collagenous NC4 #status predicted <NC4>  
 F:25-161/Domain: collagenous COL3 #status predicted <CO1>  
 F:162-176/Domain: non-collagenous NC3 #status predicted <NC3>  
 F:177-515/Domain: collagenous COL2 #status predicted <CO2>  
 F:516-546/Domain: non-collagenous NC2 #status predicted <NC2>  
 F:547-657/Domain: collagenous COL1 #status predicted <CO1>  
 F:658-675/Domain: non-collagenous NC1 #status predicted <NC1>  
 F:122/Modified site: pyrocollidone carboxylic acid (Gln) (in mature form) #status predicte  
 F:137-143-146-149-152-155-179-182-185-552-558/Modified site: hydroxyproline (Pro) #stat  
 F:170-174-525-658-663/Disulfide bonds: interchain #status predicted

Query Match 7.9%; Score 146; DB 2; Length 675;  
 Best Local Similarity 28.8%; Pred. No. 0.0089;  
 Matches 64; Conservative 9; Mismatches 71; Indels 78; Gaps 12;

QY 166 RVGSPRGAVPRKPEVSEMPMERDRGAHSLPEKENTLPDPTSNATSRGAAGLPPSP 225  
 DB 23 RVG--PQPPGPPGPSPGSKGIDG-----EPGSLGPP-----GKXGAPK 65

QY 226 PAVVAGG--GLALLLGVAGAGCMWRERAPPSRRPG-----PSPFRG- 272  
 DB 66 PGAGBAGLPLP-----GVDLTGT-----DGPFGNPGPDGALGPAGPPPAKGL 115

QY 273 -----GSLGCGGGM-----GPR-----EAFEGELIALRG----- 300

DB 116 PGPPGPPGSPSLPGNPNFRGPPGSPGLPGFPGPFGPGLAGLIPGGGDIQCPALCP 175

QY 301 -GAADPPCFPEYKVSQDYGHFVIVODGP-----PQSPENI 336  
 DB 176 PGPPGPPGMPFGKHTGKHKBPGELTKEGKSGRPPPGI 217

## RESULT 26

hypothetical protein F14F7.1 - *Caenorhabditis elegans*

C/Species: *Caenorhabditis elegans*  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000

C/Accession: T20906  
R/Murray, A.  
submitted to the EMBL Data Library, November 1996

A/Reference number: Z19345

A/Accession: T20906

A/Status: Preliminary; translated from GB/EMBL/DBD

A/Molecule type: DNA

A/Residues: 1-305 <MW>

A/Cross-references: EMBL:Z81503; PDB:CA80411.1; GSPDB:GN00021; CESP:F14F7.1

A/Experimental source: clone F14F7

C/Genetics:

A/Map position: 3

A/Insertions: 27/3; 49/3

C/Superfamily: unassigned collagens

Query Match 7.9%; Score 145.5; DB 2; Length 305;

Best Local Similarity 22.5%; Pred. No. 0.0041;

Matches 83; Conservative 21; Mismatches 122; Indels 143; Gaps 12;

```

QY 11 VVVGALLLG-----VLGLVSGLSLEPPVWN-----SANK----- 40
DB 5 VLLKAVRLAFSAVAFVCLVSVVCITLPMYNYVSGKRVLVGEVSPCKSAHEVFTEV 64
QY 41 -----RFOAGGVLYXPOIGDRDLDCPRARPPGHSNNYFYLYLVGG 86
DB 65 NPLRASATNSTSAHAGVGY-----AQPGGGGG-----GG 97
QY 87 AGRCCEAPPAENLLTCDRPDLRFTIKFQEVSPNLMGHEFRSHDYIATSDTRE 146
DB 98 GGCSTCCRCRPGAGTTPKPG-----RPGAPGAMG----- 128
QY 147 GLESLQGVCLTRGKVKVLRVQSGPRGAVPRKPYSEMPMERDRGAHSLPCKENLPD 206
DB 129 -----MPGNPKGSGGPGHPVPTPKCPGPG-----RPGPPGPPGR 165
QY 207 PTSN-----ATSRGA-----EGPLPPSPMAVAGAGLALLLVAGAGAMCM 251
DB 166 PSSDQPGPRPATGGAAPRPPGPKRGARAGNSGRAGAPQPDNDHAGCGVGARPP 225
QY 252 RRRRAKPSRRPPGSGFGSGSLGCGGGGN-----GPREAPGSLGIALRGGAADPP 307
DB 226 AGPRGAPGAGHGGSSGGGRPGAPGKAPGQPGRRPP-DGHPGQPGRRPGSGSGNRGV 284
QY 308 CPHYEKVSG 316
DB 285 CPKYCAIDG 293

```

## RESULT 27

collagen alpha 1(X) chain precursor - bovine

C/Species: *Bos primigenius taurus* (cattle)

C/Date: 21-Nov-1993 #sequence\_revision 23-Feb-1996 #text\_change 13-Aug-1999

C/Accession: S13301

R/Thomas, J. T.; Kwan, A. P. L.; Grant, M. E.; Boot-Handford, R. P.

Biochem. J. 273, 141-148, 1991

A/Title: Isolation of cDNAs encoding the complete sequence of bovine type X collagen. EV

A/Reference number: S13301; MUID:9111331; PMID:1703407

A/Accession: S13301

A/Molecule type: mRNA

A/Residues: 1-674 <THO>

A/Cross-references: EMBL:X53556; NID:9263; PDB:CAA37624.1; PID:9264

C/Genetics:

A/Genes: COL10A1

C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

C/Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer

F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-674/Product: collagen alpha 1(X) chain #status predicted <MAT>  
F:547-673/Domain: complement C1q carboxyl-terminal homology <CIQ>

Query Match 7.8%; Score 145; DB 2; Length 674;

Best Local Similarity 24.2%; Pred. No. 0.01;

Matches 100; Conservative 30; Mismatches 125; Indels 158; Gaps 24;

```

QY 15 ALLILGVLSGLSLDEPPYNNANSKRFOAGGVLYYPQIGRDLDCRA----- 65
DB 6 ALLILMSLVLRG-----VFP--TERVQTPG-IKGPSNKTQFTIYALKGKVSIR 56
QY 66 -----RPPGP-----HSSPNYEFYLYLVG-----GAQR-RCEADPAPNLLIT 103
DB 57 GEOGIPGPPGAPRGRHPSPS-----GPPKRGTSPPQCPGLPGPPGS---A 104
QY 104 CDRPDLRFTIKFQEVSNLMGHEFRSHDYIATSDTEGSLG----- 153
DB 105 TGKPGI-----DGLPEKQ-----GER-GLNGPKDIDGPAGLPGR 137
QY 154 -----GVCLTRGKVKVLRVQ-SPPRGAVPR-----KPVSEMP-----MERDRGAHSL 196
DB 138 RGPGRPPGIPGPAGISVPGKPGQPTGPRGRPPGKGTSGVGLNGKGMGHTCTC 197
QY 197 EPCKENLQEDPPTSNATSRGABGLPPP-----SMPAVAGAAGLALLLVAGAGGA 248
DB 198 RPERGLPG-----PQPTGPPGPPGVGKRGKENGCLPQCPKDKD-----QGVGERGA 245
QY 249 MCRRRAKPSRRPP-----GPGSFGSGSLGCGGGGPREAPGELGIA----- 296
DB 246 -----AGPSGQSPGPGQSGEGIGKPPAPGIPGPKKGPGAPGAGTGLGAPGF 298
QY 297 -----LRG-GGADPPPCPHYEKVSGDYHPVYIVDGP-----PQSPENI 336
DB 299 GKPELPLGLKGQGRPVGLPSPGAKGEGPAGHGEAGLPSPGNNMPQGPKEI 351

```

## RESULT 28

collagen alpha 1(IX) chain precursor, long splice form - human

N/Alternate names: procollagen alpha 1(IX) chain, long splice form

C/Species: *Homo sapiens* (man)

C/Date: 13-Jan-1995 #sequence\_revision 10-Feb-1995 #text\_change 20-Sep-1999

C/Accession: S13580; S23295; S02140; G35980; S21887; S74294

R/Muragaki, Y.; Kimura, T.; Ninomiya, Y.; Olsen, B. R.

Eur. J. Biochem. 192, 703-708, 1990

A/Title: The complete primary structure of two distinct forms of human alpha-1(IX) col1

A/Reference number: S13580; MUID:91006164; PMID:2209617

A/Accession: S13580

A/Molecule type: mRNA

A/Residues: 1-931 <MR>

A/Cross-references: EMBL:X54412; NID:930083; PDB:CA38276.1; PID:930084

A/Residues: 580-596,598-620,'R',622-813,835-884 <XIN>

R/Ninomiya, Y.; Caesteegnols, F.; Gercke, D.; Gordon, M. K.; Jacenko, O.; Luvalle, P.; Mc

maguchi, N.; Olsen, B. R.

in Extracellular Matrix Genes, Sandell L. J. and Boyd C. D., eds., pp.79-114, Academic P

A/Title: The molecular biology of collagens with short triple-helical domains.

A/Reference number: S22243

A/Accession: S23295

A/Status: not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 1-253,'V',255-815,835-884 <XIN>

R/Kimura, T.; Mattei, M. G.; Stevens, J. W.; Goldring, M. B.; Ninomiya, Y.; Olsen, B. R.

Eur. J. Biochem. 179, 71-78, 1989

A/Title: Molecular cloning of rat and human type IX collagen cDNA and localization of t

A/Reference number: S02140; MUID:89137096; PMID:2465149

A/Accession: S02140

A/Status: not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 580-596,598-620,'R',622-813,835-884 <XIN>

R/Muragaki, Y.; Ninomiya, Y.; Henney, A.; Ninomiya, Y.; Olsen, B. R.

Proc. Natl. Acad. Sci. U.S.A. 87, 2400-2404, 1990

A/Title: The alpha1(IX) collagen gene gives rise to two different transcripts in both m

A/Reference number: A35980; MUID:30207204; PMID:1690886

A/Accession: C35980  
 A/Molecule type: DNA  
 A/Residues: 1-4:233-267 <MUR2>  
 A/Cross-references: EMBL:M32135  
 R/Olsen, B.R.  
 Submitted to the EMBL Data Library, February 1990  
 A/Reference number: S21087  
 A/Accession: S21087  
 A/Molecule type: DNA  
 A/Residues: 1-4:233-248, 'N', 250-267 <OLS>  
 A/Cross-references: EMBL:M32135  
 R/Diab, M.; Wu, J.U.; Byre, D.R.  
 Biochem. J. 314, 327-332, 1996  
 A/Title: Collagen type IX from human cartilage: a structural profile of intermolecular  
 A/Reference number: S64673; MUID:96195147; PMID:8660302  
 A/Accession: S74294  
 A/Molecule type: protein  
 A/Residues: 405-417 <DIA>  
 C/Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
 ed and subsequently O-glycosylated.  
 C/Genetics:  
 A/Genes: GDB:COL9A1  
 A/Cross-references: GDB:119794; OMIM:120210  
 A/Map position: 6q12-6q14  
 A/Intons: 5/2; 232/3; 260/3; 267/3  
 C/Function:  
 A/Description: structure of extracellular fibrous polymer associated with type  
 A/Note: in chondrocytes the long splice form is predominantly produced  
 C/Superfamily: unassigned collagens  
 C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; heter  
 F/1-23/Domain: signal sequence #status predicted <SIG>  
 F/24-331/Domain: collagen alpha 1(I) chain, long splice form #status predicted <MAT>  
 F/24-268/Domain: non-collagenous NC4 #status predicted <NC4>  
 F/269-405/Domain: collagenous COL3 #status predicted <COL3>  
 F/406-417/Domain: non-collagenous NC3 #status predicted <NC3>  
 F/418-756/Domain: collagenous COL2 #status predicted <COL2>  
 F/757-786/Domain: non-collagenous NC2 #status predicted <NC2>  
 F/787-901/Domain: collagenous COL1 #status predicted <COL1>  
 F/902-931/Domain: non-collagenous NC1 #status predicted <NC1>  
 F/171/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 7.8%; Score 145; DB 2; Length 931;  
 Best Local Similarity 22.3%; Pred. No. 0.015;  
 Matches 84; Conservative 21; Mismatches 120; Indels 152; Gaps 13;

QY 2 GPHSPGSGVAVGALLLGLVGLSLSPYVWNSANKRFQAGGYLVLPQIGRLDL 61  
 DB 391 GPP--GPPGR-----GTIGFHDG-----DPL 410

QY 62 CPAPRPPGPHSPNYEFYKLVG---GAQGRCEAPPAENLLTCRPLDLRPTIK 116  
 DB 411 CPAPRPPGPRGPG---LPGRGHHKAGKEIGE----- 440

QY 117 FGYSYPLMGEHFRSHHDYITATSDTRGSLQGVCLTRGMKVLTVGSGPRGAV 176  
 DB 441 -----PGRQH-----KGEKDG-----ELSEVGAQGPRAQ 468

QY 177 PRKPEEMPERDGAHSL--EPGKENTPGDPTSNAT-----SRGAEGLPPP 223  
 DB 469 GLRGITGLVDEKGEKARGLDGEPFGQJFGAGDQGRGPRGAPKDRGARGARIP 528

QY 224 SMRAVAGAAAGLLLLLVAGAGACWRRRAKPSSESHNG-----PSFGSGSISG 276  
 DB 529 GLRGPKDNTD---LPGVGRDGTGMPGTGEPGKPGPPGAGIQLGVPGLPFGAG 583

QY 277 LGGGGSGVPRPAGELGIALRG---GAADPPCPHYEKVSG 316  
 DB 584 VAEKSGTAPRGQMGNSGRPGQGGPRGPRGPGQLPGRSGELGVGPGFLPGRKL 643

QY 317 DYGHPIYVDGPPQSP 333  
 DB 644 SLGSPGLPGLPGRPGP 660

RESULT 29  
 553787  
 collagen alpha chain - Paralvinella grasslei (fragments)  
 C/Species: Paralvinella grasslei  
 C/Date: 19-Mar-1997 #sequence\_revision 18-Jul-1997 #text\_change 07-May-1999  
 C/Accession: S53787  
 R/Gall, F.; Mann, K.; Wiedemann, H.; Engel, J.; Timpl, R.  
 J. Mol. Biol. 246, 284-294, 1995  
 A/Title: Structural comparison of cuticle and interstitial collagens from annelids livi  
 A/Reference number: S53786; MUID:95173973; PMID:7869380  
 A/Accession: S53787  
 A/Molecule type: protein  
 A/Residues: 1-90:91-254:255-304:305-374:375-438 <GAI>  
 C/Superfamily: unassigned collagens  
 C/Keywords: coiled coil; extracellular matrix; glycoprotein

Query Match 7.8%; Score 144.5; DB 2; Length 438;  
 Best Local Similarity 26.5%; Pred. No. 0.0071;  
 Matches 101; Conservative 15; Mismatches 124; Indels 141; Gaps 22;

QY 2 GPP-----HSPGSGVAVGALLLGLVGLS-----GLSLSPYVWNSANKRFQ 43  
 DB 105 GPPGLTSSXGPAG--PXGANGLPQTQFAGAGPXPSPGPRGRLGHP---GEAGRGQ 160

QY 44 AEGGYLVLPQ-----IGRLDLCPRAPP-----GPHSPNYEFYKLVG 85  
 DB 161 RGVGALGFPPOGRGLPGR---GLPGRGPRGAGRGGRGPPSP-----G 207

QY 86 GAQG--RCAPPAENLLTCRPLDLRPTIKFQEYSPLMGEHFRSHHDYITATSDG 143  
 DB 208 GPQDSDGAGTPGAPGARGAPRG-----SDG 235

QY 144 TREGLSLQ-----GYCLTRGMKVLTVGSGPRGAVPRKPVSEMERDGAHSLP 198  
 DB 236 -REGKPPSSGKPGKDG-----RPDGRG-----LPGRGAAGS--N 269

QY 199 GKENTPGDPTSNATSRGAEGLPSPMPAVAGA--GLALLLVAGAGACWRRRA 256  
 DB 270 GEDQNGKP-----GKGVGPPGIQLPGLPGEARGKELQGLDGERGK---PGERG 319

QY 257 KPSRRPPG-----PGRFRGSLGSGGGMGPRAEPGELIARSGGAADPPCPHYE 312  
 DB 320 KDGSGRHPTAGLPFGPGPGV--KGERGDHGR--GPGSPG---SXGERGPGKXPQ 372

QY 313 KYSGDYGHPIYVDGPPQSP 333  
 DB 373 GLRGDRNTGREGPRGAPGP 393

RESULT 30  
 528774  
 collagen alpha chain - tube worm (Riftia pachytilia) (fragment)  
 C/Species: Riftia pachytilia  
 C/Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 23-May-1997  
 C/Accession: S28774; S22915; S17581  
 R/Mann, K.; Gall, F.; Timpl, R.  
 Eur. J. Biochem. 210, 839-847, 1992  
 A/Title: Amino-acid sequence and cell-adhesion activity of a fibril-forming collagen fr  
 A/Reference number: S28774; MUID:93110909; PMID:1483468  
 A/Accession: S28774  
 A/Molecule type: protein  
 A/Residues: 1-95, 'X', 97-107, 'X', 109-191, 'X', 193-260, 'X', 262-278, 'X', 280-572, 'X', 574-611,  
 A/Note: we have shown the unidentified residues as lys forming glycosylated 5-hydroxylys  
 R/Mann, K.; Gall, F.; Timpl, R.  
 submitted to the Protein Sequence Database, July 1992  
 A/Description: Amino acid sequence and cell adhesion activity of a fibril-forming collag  
 A/Reference number: S22915  
 A/Accession: S22915  
 A/Molecule type: protein  
 A/Residues: 1-95, 'X', 97-107, 'X', 109-191, 'X', 193-260, 'X', 262-278, 'X', 280-572, 'X', 574-611,  
 A/Note: 903-proline modified to 4-hydroxyproline was also found

R;Gall, F.; Wiedemann, H.; Mann, K.; Kuehn, K.; Timpl, R.; Engel, J.  
 J. Mol. Biol. 221, 209-223, 1991  
 A:Title: Molecular characterization of cuticle and interstitial collagens from worms COL  
 A:Reference number: S17581, MUID:92015209, PMID:1920405  
 A:Accession: S17581  
 A:Molecule type: Protein  
 A:Residues: 8-45;525-545, 'X', 547-566, 'X', 568-572, 'X', 574-611, 'X', 613-618, 'X', 811-882 <CA  
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (h  
 h 3- and 4-hydroxylated in the X-position. Lysines are 5-hydroxylated and subsequently a  
 C:Complex: homotrimer  
 C:Superfamily: unassigned collagens  
 C:Keywords: cell binding; coiled coil; extracellular matrix; glycoprotein; homotrimer; h  
 F:1-12/Domain: amino-terminal telopeptide (fragment) <NTE>  
 F:113-1023/Domain: collagenous #status experimental <COL>  
 F:1024-1027/Domain: carboxyl-terminal telopeptide (fragment) <CTE>  
 F:21,24,123,243,273,276,285,291,303,348,381,621,645/Modified site: 4-hydroxyproline (Pro  
 F:217,39,54,72,90,93,128,150,162,165,174,177,180,207,216,229,228,237,249,255,306,312,321,  
 F:711,714,717,723,744,759,774,783,792,816,843,849,855,861,867,888,894,915,945,954,963,966  
 F:753,161,165,416,551,647,773,815,1010,1013,1016,1019/Modified site: 3-hydroxyproline (P  
 F:196,108,192,261,279,573,612,657,738,765,810,927,936/Modified site: 5-hydroxylysine (Lys  
 F:166,108,192,261,279,573,612,657,738,765,810,927,936/Binding site: carboxylate (Lys) (C  
 F:183,342,546,567,939/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F:351,933/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental

Query Match 7.8%; Score 144; DB 2; Length 1027;  
 Best Local Similarity 22.2%; Pred. No. 0.019;  
 Matches 82; Conservative 24; Mismatches 113; Indels 150; Gaps 17;

46 GGVVLYPGIGRDLDCPRARPPGPHSSPNVEFYKLVVGAQGRCS----- 93  
 4 GRRYDAQVGP-----IGRRG-PPGPPSPGQO-----GYGLRGPPGSGMGPRGK 50  
 94 -APPANLLTCDRDLRFTIKFOEYSPNLMGHEFPHSHDYIATSDTREGLESIQ 152  
 51 RPPGP-----AGIAG-----KSGDGRDGPBP 75  
 153 GGVCVT-----RGMYLV-----LRVQSPRG--AVPKRPSVM 184  
 76 GIGVMPGPGAGMGMPGPKRFRGLSGSKGQSGSNQVGGPGPGPBGPIPR 135  
 185 PWERDRGAHSLPEKKNLPQDPTSNATSRGAEPLPPSPAVAGAAAGLAL----- 238  
 136 GQTERG-----RDGKSGLP-----LRGVGLAGFPFPGPG-IGTSGSPGFPPTGSK 183  
 239 -----LLGVAGAGAMCMRRRAKPSSESRHPG-PSFSGGSGSLGCGGGM----- 283  
 184 GDRGSGTIGAGQGLQPGVGLSGQPGVAGBNHGMGMGMDGANGEBGASGESGLPPSGFP 243  
 284 GPR-----EABPGLGIALRG-----GADPPCPHYEKVSGDYGHPIYI 324  
 244 GPRGMPTGSGSCQAGAKQDGGFTGEQGRPGAPGVKSSGPP-----GDVGAPGHA 294  
 325 VODGPPSP 333  
 295 GEAGKRGSP 303

## RESULT 31

A54121  
 collagen alpha-4 chain precursor - sea urchin (strongylocentrotus purpuratus)  
 N:Alternate names: collagen alpha 2(IV) chain homolog  
 C:Species: Strongylocentrotus purpuratus (purple urchin)  
 C:Date: 07-Jul-1995 #sequence\_revision 07-Jul-1995 #text\_change 13-Aug-1999  
 C:Accession: A54121, S4317  
 R:Exposito, J.Y.; Suzuki, H.; Geoufjon, C.; Garrone, R.; Solursh, M.; Ramirez, F.  
 J. Biol. Chem. 269, 13167-13171, 1994  
 A:Title: Identification of a cell lineage-specific gene coding for a sea urchin alpha2(I  
 A:Reference number: A54121, MUID:94230414, PMID:8175744  
 A:Accession: A54121  
 A:Molecule type: mRNA  
 A:Residues: 1-1747 <EXP>  
 A:Cross-references: EMBL:X76730; NID:9483606; PIDN:CAA4146.1; PID:9483607  
 C:Genetics:

A:Gene: COL4alpha  
 C:Superfamily: collagen alpha 1(IV) chain

Query Match 7.7%; Score 143; DB 2; Length 1747;  
 Best Local Similarity 24.1%; Pred. No. 0.04;  
 Matches 84; Conservative 24; Mismatches 159; Indels 82; Gaps 14;

7 GPGVAVGALLLGLVGLVSGLSLEPVYNSANKRFOAGGVLYVQIDRLDCPRAR 66  
 266 GPRGMG-----MGATGEVGLDSYDYEKGLPGVSGRGPFGNGITGLGMNEKER 321  
 67 PPGPHSSPNVEFYKLVVGAQGRCPADPAPNLLTCDRDLRFTIKFOEYSPNLMG 126  
 322 -DPPQRPQYDGRK--GPGSDYGMDDPDPAPDI----- 352  
 127 HERSHHDYIATSDTREGLESLOGVCLRMKLVLRVGO-----SPRGAVPRKPS 182  
 353 -----EETVGVGPDEGPPGNDFPRGSGALGTFDQGPYVGMGPRGPRGTG 405  
 183 EMERDRGAHSLPEKKNLPQD-----TSNATSRGAEPL-----PPSPAVAGAA 232  
 406 SQEGSDGEKSGRGIDPGFOETGERGNDGNFGRGKMGVGRBPFGMDSGRGQ 465  
 233 GGLALLLVGAAGAMCMRRRAKPSSESRHPG-----PSFRGSGSLGCGGGMGP 285  
 466 G-----FMQCKGRGP-----PGRAGPAGPARSGNOSFGFPDDTGNLTGKMGIGQALG 516  
 286 REAPPELGIALRGGAADPPCPHYEK-VSGDYGHPIYVODGPPSP 333  
 517 RDRPFGSKG-----ELGGIC--PFCPPGKGVPPDRGV-----GPDGSP 554

## RESULT 32

S23779  
 collagen alpha 1(VIII) chain - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
 C:Accession: S23779  
 R:Muragaki, Y.; Shiota, C.; Inoue, M.; Ooshima, A.; Olsen, B.R.; Nimomiya, Y.  
 Eur. J. Biochem. 207, 895-902, 1992  
 A:Title: Alpha-1(VIII)-collagen gene transcripts encode a short-chain collagen polypept  
 A:Reference number: S23779; MUID:92362626; PMID:1499564  
 A:Accession: S23779  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-743 <MUR>  
 A:Cross-references: EMBL:X66976; NID:950493; PIDN:CAA47387.1; PID:91359953  
 C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
 F:616-742/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 7.7%; Score 142.5; DB 1; Length 743;  
 Best Local Similarity 21.3%; Pred. No. 0.017; Mismatches 136; Indels 209; Gaps 21;

14 GALLLGLVGLVSGLSLEPVYNSANKRFOAGGVY---LYPQIGDRDLDCPRARPG 69  
 6 GRLQLGLIFIT-----LNSV--RLIQAGAYYGRKLPQIPQIPQIPQIPQIPQ 55  
 70 ---PH-----SSPNVEFYLY-----L 83  
 56 QQVPHMPLGKGLSKGEMPHQYGEYVLPQMKELPVPVMGKEVPPKGRKVPPLAS 115  
 84 VGAQGRGRC---APPANLLTCDRDLRFTIKFOEYSPNLMGHEFRSHHDYIITA 139  
 116 LRGEQPRGPRGPRGP-----PGLPGHMPG-----IK 146  
 140 TSDGT-----REGLESIG--GVCLTGMVYLVRVQSPRGGAVERKPEVMEPRER 190  
 147 GKPDPQYDGIQPGMGMGKRGAMGAK-----GEIGPKGEIGPMGIPGQ 196  
 191 G--AAISL-----EPKKNLPQDPTSNATSRGAEGLPPSPM----- 225  
 197 GPPGPRGLPGIGRPGRGLPGQPGAKG-ERGPKRPPGPRGLGPRKGEKGFGLGLG 255



QY 226 -PAVAGAGGATLALLGVAGAGMCMRRRRAPSPSRHPG----- 265  
 Db 256 PPGMHGPPGPGVPGVGVKPGVTPGPGQGLGKRPGEPRGPGQLGVPGVGPFGMPGV 315  
 QY 266 --PSPFRGSLGIGGGGG-----MGPRE 287  
 Db 316 GKPPQDSSRRQPGPGGKGGQGLPGLPGPLDGVGKPGPGGKRGHGGVGVGPR- 374  
 QY 288 AEPGELIALRGGADDPFCFHYEKVSDYGHVTVQDP-----PGSP 334  
 Db 375 GKXGPIAPMGG-----PGEFGLPGIIPGMGPGALGFGPGEGGVGPGQPP 425

RESULT 33  
 CGH2V  
 collagen alpha 2(V) chain precursor - human  
 C/Spectes: Homo sapiens (man)  
 C/Date: 31-Jul-1989 #sequence revision 28-Jul-1995 #text change 31-Dec-2000  
 C/Accession: A31427; A54555; S43643; A25874; I55239; I59025; A25374; A30017  
 R/Woodbury, D.; Benson-Chanda, V.; Ramirez, F.  
 J. Biol. Chem. 264, 2735-2738, 1989  
 A>Title: Amino-terminal propeptide of human pro-alpha2(V) collagen conforms to the struc  
 A/Reference number: A31427; MUID:89123368; PMID:2914927  
 A/Accession: A31427  
 A/Molecule type: mRNA  
 A/Residues: 1-463 <MO>  
 A/Cross-references: GB:J04478; NID:q179697; PIDN:AAAS1859.1; PID:q179698  
 R/Experimental source: Placenta  
 R/Greenspan, D.S.; Lee, S.T.; Lee, B.S.; Hoffman, G.G.  
 Gene Expr. 1, 29-39, 1991  
 A>Title: Homology between alpha2(V) and alpha1(III) collagen promoters and evidence for  
 A/Reference number: A54555; MUID:92314691; PMID:1820205  
 A/Accession: A54555  
 A/Molecule type: DNA  
 A/Residues: 1-32 <GRE>  
 A/Cross-references: GB:M58529; NID:q180634; PIDN:AAAC1699.1; PID:5553235  
 R/Moradi-Neeli, M.; Rousseau, J.C.; Kleman, J.P.; Champliand, M.F.; Boucillon, M.M.; Ben  
 Eur. J. Biochem. 221, 987-995, 1994  
 A>Title: Diversity in the processing events at the N-terminus of type-V collagen.  
 A/Reference number: S43642; MUID:94237164; PMID:8181482  
 A/Accession: S43643  
 A/Molecule type: Protein  
 A/Residues: 288-291, 'P', 293-294, 'X', 296-297, 606, 'X', 608-617 <MOR>  
 R/Weil, D.; Bernard, M.; Garcano, S.; Ramirez, F.  
 Nucleic Acids Res. 15, 181-189, 1987  
 A>Title: The pro alpha 2(V) collagen gene is evolutionarily related to the major fibrill  
 A/Reference number: A25874; MUID:87146331; PMID:3029669  
 A/Accession: A25874  
 A/Molecule type: mRNA; DNA  
 A/Residues: 398-1496 <WEI>  
 A/Cross-references: GB:X04558; NID:q29588; PIDN:CAA2454.1; PID:q1340175  
 R/Experimental source: Rhabdomyosarcoma cell line  
 R/Wyers, U.C.; Loidl, H.R.; Scollie, C.A.; Seyer, J.M.  
 J. Biol. Chem. 260, 5533-5541, 1985  
 A>Title: Partial covalent structure of the human alpha 2 type V collagen chain.  
 A/Reference number: I55239; MUID:85182703; PMID:2985598  
 A/Accession: I55239  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1002-1226 <RE2>  
 A/Cross-references: GB:M10956; NID:q180427; PIDN:AAAS2007.1; PID:q180428  
 A/Note: part of this sequence were determined by protein sequencing  
 R/Manuel, B.S.; Camizazaro, L.A.; Seyer, J.M.; Wyers, J.C.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985  
 A>Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm  
 A/Reference number: I59025; MUID:85216505; PMID:3858626  
 A/Accession: I59025  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1003-1034 <RES>  
 A/Cross-references: GB:M1135; NID:q179693; PIDN:AAAS1857.1; PID:q179694  
 A/Note: part of this sequence were determined by protein sequencing

R/Wyers, J.C.; Loidl, H.R.; Seyer, J.M.; Dion, A.S.  
 J. Biol. Chem. 260, 11216-11222, 1985  
 A>Title: Complete primary structure of the human alpha-2 type V procollagen COOH-termin  
 A/Reference number: A25374; MUID:85289337; PMID:2411731  
 A/Accession: A25374  
 A/Molecule type: mRNA  
 A/Residues: 1227-1417, 'T', 1419-1437, 'S', 1439-1496 <MYE>  
 A/Cross-references: GB:M11718; NID:q180912; PIDN:AAAS2058.1; PID:q180913  
 R/Experimental source: normal fibroblasts  
 R/Tsipouras, P.; Schwartz, R.C.; Liddell, A.C.; Salkeid, C.S.; Weil, D.; Ramirez, F.  
 Genomics 3, 275-277, 1988  
 A>Title: Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2, located on  
 A/Reference number: A30017; MUID:89138450; PMID:3224983  
 A/Accession: A30017  
 A/Molecule type: DNA  
 A/Residues: 1449-1463, 'E', 1465-1495, 'A' <TSI>  
 A/Cross-references: GB:J03051; NID:q179695; PIDN:AAAS1858.1; PID:q179696  
 A/Note: the authors translated the codon GAA for residue 1460 as Gln, and GAG for resid  
 C/Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
 are 5-hydroxylated and subsequently O-glycosylated.  
 C/Comment: The amino-terminal propeptide domain appears not to be completely cleaved.  
 C/Genetics:  
 A/Gene: GDB:COL5A2  
 A/Cross-references: GDB:119064; OMIM:120190  
 A/Map position: 2q31-2q31  
 A/Intons: 33/1; 812/3; 830/3; 848/3; 884/3; 902/3; 922/3; 974/3; 1046/3; 1064/3; 1448/  
 C/Complex: type V collagen may be a homotrimer of alpha 1(V) chains (see PIR:CGH01V), a  
 alpha 2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among th  
 length, is formed with desmosine cross-links made from lysine and allysine residues  
 C/Function:  
 A/Description: structural component of extracellular fibrous polymer associated with ce  
 A/Note: may play a role in controlling the lateral growth of collagen I fibrils  
 C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology  
 C/Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyprol  
 P.1-26/Domain: signal sequence #status predicted <SIG>  
 P.127-1250/Product: collagen alpha 2(V) chain #status predicted <MAT>  
 P.127-193/Domain: amino-terminal propeptide (uncleaved) #status predicted <NPP>  
 P.127-108/Region: nonhelical  
 P.140-99/Domain: von Willebrand factor type C repeat homology <WVC>  
 P.109-166/Region: helical  
 P.187-208/Region: nonhelical  
 P.209-1225/Region: helical  
 P.503-505/Region: cell attachment (R-G-D) motif  
 P.591-943/Region: cell attachment (R-G-D) motif  
 P.1064-1066/Region: cell attachment (R-G-D) motif  
 P.1067-1068/Region: cell attachment (R-G-D) motif  
 P.1097-1099/Region: cell attachment (R-G-D) motif  
 P.1124-1126/Region: cell attachment (R-G-D) motif  
 P.1123-1135/Region: cell attachment (R-G-D) motif  
 P.1125-1250/Region: carboxyl-terminal nonhelical telopeptide  
 P.1251-1496/Domain: fibrillar collagen carboxyl-terminal homology <FCC>  
 P.127/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicte  
 P.193-194/Cleavage site: Ala-Gln (procollagen N-endopeptidase) #status predicted  
 P.194/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predict  
 P.201/Modified site: allysine (lys) #status predicted  
 P.290, 293, 296, 608, 614, 1004, 1007, 1013, 1028, 1034/Modified site: 4-hydroxyproline (Pro) #s  
 P.299, 1139/Modified site: 5-hydroxylysine (lys) #status predicted  
 P.299, 1139/Binding site: carboxylate (lys) (covalent) #status predicted  
 P.1025/Modified site: 5-hydroxylysine (lys) #status experimental  
 P.1250-1251/Cleavage site: Glu-Asp (procollagen C-endopeptidase) #status predicted  
 P.1259, 1397/Binding site: carboxylate (Asn) (covalent) #status predicted  
 P.1253, 1299, 1325/Disulfide bonds: interchain #status predicted  
 P.1333-1494, 1402-1447/Disulfide bonds: #status predicted

Query Match 7.7%; Score 142.5; DB 1; Length 1496;  
 Best Local Similarity 29.1%; Pred. No. 0.037;  
 Matches 55; Conservative 15; Mismatches 76; Indels 43; Gaps 9;

QY 169 QSPRG--GAV-FRXPVSEMPERDRGAHSLPEKENVLP-----DPTSNATSRGAE 217  
 Db 500 RGRGDDPRTLLPDPVBERGAPGNRGRF-----PSSDGLPGRKAGRGPRGVSSGPKQSG 554

```

QY 218 GELPPSPNPAVAGAGLALLLGVAGAGAMCWRRRRAKPSRRHPGP-----SFGRCG 273
DB 555 GDRPGRGPGULPGARG-----LTGNPVGQPEKGLGPGAGDGRGPPPSITIKQPG 609
QY 274 SLGLGGGGGMBREARPEBELGIALRGGAADPPFCPHYEKVSGDYGHVYVQDGP-- 330
DB 610 TMLGPGPGKSGNDPGKPGKPEAG---NPGVPQRGAPGKDGHYGVYGP-----GPGQLR 659
QY 331 -----QSPG 334
DB 660 GERGEQGP 668

RESULT 34
S15435
collagen alpha 1(VIII) chain precursor - human
C/Species: Homo sapiens (man)
C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 13-Aug-1999
C/Accession: S15435
R/Muragaki, Y.; Mattei, M.G.; Yamaguchi, N.; Olsen, B.R.; Minomiy, Y.
Eur. J. Biochem. 197, 615-622, 1991
A/Title: The complete primary structure of the human alpha-1(VIII) chain and assignment
A/Reference number: S15435; MUID:91231001; PMID:2029894
A/Accession: S15435
A/Status: Preliminary
A/Molecule type: mRNA
A/Residues: 1-744 <MUR>
A/Cross-references: EMBL:X57527; NID:G30081; PIDN:CAA40748.1; PID:G30082
A/Genes: GDB:COL8A1
A/Cross-references: GDB:128104; OMTM:120251
A/Map position: 3q11.1-3q13.2
C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology
F/1-20/Domain: signal sequence #status predicted <SIG>
F/21-744/Product: collagen alpha 1(VIII) chain #status predicted <MAT>
F/21-117/Region: amino-terminal nonhelical
F/118-571/Region: interrupted helical
F/572-744/Region: carboxyl-terminal nonhelical
F/617-743/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 7.7%; Score 142; DB 2; Length 744;
Best Local Similarity 23.5%; Pred. No. 0.019;
Matches 73; Conservative 16; Mismatches 96; Indels 126; Gaps 14;

```

```

C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
C/Accession: S21369; S28808; S13745; S18663
R/Brabant, X.Y.Z.; Bardon, X.Y.Z.; Dani, C.
Submitted to the EMBL Data Library, April 1992
A/Reference number: S21369
A/Accession: S21369
A/Molecule type: mRNA
A/Residues: 1-1029 <IBR>
A/Cross-references: EMBL:X65582; NID:G49808; PIDN:CAA46541.1; PID:G49809
R/Brabant, A.; Bertrand, B.; Bardon, S.; Amri, E.Z.; Grimaldi, P.; Allaud, G.; Dani,
Biochem. J. 289, 141-147, 1993
A/Title: Cloning of alpha2 chain of type VI collagen and expression during mouse devel
A/Reference number: S28808; MUID:93143659; PMID:8380980
A/Accession: S28808
A/Molecule type: mRNA
A/Residues: 266-1029 <IB2>
A/Cross-references: EMBL:X62332; NID:G49906; PIDN:CAA44206.1; PID:G49907
R/Constantinou, C.D.; Ulmen, S.A.
Matrix 11, 1-9, 1991
A/Title: Structure of CDNA encoding the triple-helical domain of murine alpha-2(VI) c
rate oligonucleotides for generation of novel CDNA clones.
A/Reference number: S13745; MUID:91226374; PMID:1709252
A/Accession: S13745
A/Molecule type: mRNA
A/Residues: 266-267; S', 269-294; 'L', 296-600 <CON>
A/Cross-references: GB:I06343; NID:G192671; PIDN:AAA37441.1; PID:G192672
A/Note: the sequence from Fig. 3 is inconsistent with that from Fig. 4 in having 306-I;
C/Superfamily: collagen alpha 1(VI) chain; von Willebrand factor type A repeat homology
C/Keywords: alternative splicing; cartilage; cell binding; connective tissue; disulfide
F/1-28/Domain: signal sequence #status predicted <SIG>
F/29-1029/Product: collagen alpha 2(VI) chain #status predicted <MAT>
F/29-265/Domain: globular #status predicted <NC2>
F/24-226/Domain: von Willebrand factor type A repeat homology <VWA1>
F/266-600/Domain: collagenous #status predicted <COL>
F/376-378/Region: cell attachment (R-G-D) motif
F/366-438/Region: cell attachment (R-G-D) motif
F/499-501/Region: cell attachment (R-G-D) motif
F/508-510/Region: cell attachment (R-G-D) motif
F/549-551/Region: cell attachment (R-G-D) motif
F/601-1029/Domain: globular #status predicted <NC1>
F/623-799/Domain: von Willebrand factor type A repeat homology <VWA2>
F/841-1012/Domain: von Willebrand factor type A repeat homology <VWA3>
F/150,337,640,907,963/Binding site: carbohydrate (Aen) (covalent) #status predicted

Query Match 7.7%; Score 142; DB 1; Length 1029;
Best Local Similarity 29.7%; Pred. No. 0.027;
Matches 90; Conservative 16; Mismatches 125; Indels 72; Gaps 20;

```



A: Molecule type: mRNA  
 A: Residues: 302-423 <CHT>  
 A: Cross-references: GB:S79877; NID:G1195576; PIDN:AA35615.1; PID:G1195577  
 R: Seyer, U.M.; Kang, A.H.  
 Biochemistry 17, 3404-3411, 1978  
 A: Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBR pe  
 A: Reference number: A90414; MUID:79000343; PMID:687591  
 A: Accession: A90414  
 A: Molecule type: Protein  
 A: Residues: 39-675, N', 677-727 <SEY3>  
 A: Experimental source: liver  
 R: Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.  
 J. Biol. Chem. 266, 5256-5259, 1991  
 A: Title: G to T transversion at position +5 of a splice donor site causes skipping of th  
 A: Reference number: 155349; MUID:91161621; PMID:1672129  
 A: Accession: 155349  
 A: Status: translated from GB/EMBL/DBJ  
 A: Molecule type: DNA  
 A: Residues: 537-605 <LEB>  
 A: Cross-references: GB:M59312; NID:G180815; PIDN:AA52041.1; PID:G180816  
 R: Seyer, U.M.; Mainardi, C.; Kang, A.H.  
 Biochemistry 19, 1583-1589, 1980  
 A: Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from ty  
 A: Reference number: A90438; MUID:80198282; PMID:6246925  
 A: Accession: A90438  
 A: Molecule type: protein  
 A: Residues: 728-895, A', 897-964 <SEY4>  
 A: Experimental source: liver  
 R: Cole, W.G.; Chido, A.A.; Lamande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Chan  
 J. Biol. Chem. 265, 17070-17077, 1990  
 A: Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping an  
 A: Reference number: A38303; MUID:91009133; PMID:2145258  
 A: Accession: A38303  
 A: Molecule type: mRNA  
 A: Residues: 861-1015 <COL>  
 A: Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:G180878; PIDN:AA55983.1; PID:G  
 A: Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syn  
 R: Manoo, B.S.; Dalgleish, R.  
 Nucleic Acids Res. 16, 2337, 1988  
 A: Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.  
 A: Reference number: S02119; MUID:88189827; PMID:3357782  
 A: Accession: S02119  
 A: Status: translation not shown  
 A: Molecule type: mRNA  
 A: Residues: 950-1018, Y', 1020-1183, S', 1185-1466 <MAN>  
 A: Cross-references: EMBL:X06700; NID:G30053; PIDN:CAA29886.1; PID:G30054  
 R: Seyer, U.M.; Kang, A.H.  
 Biochemistry 20, 2621-2627, 1981  
 A: Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from ty  
 A: Reference number: A90446; MUID:81208139; PMID:7016180  
 A: Accession: A90446  
 A: Molecule type: protein  
 A: Residues: 965-979, A', 981-984, PS', 987, QN', 990-1096, P', 1098-1152, AT', 1155, S', 1157-  
 A: Experimental source: liver  
 R: Iridi, H.R.; Brinker, J.M.; May, M.; Phlajantemi, T.; Morrow, S.; Rosenbloom, J.; Mye  
 Nucleic Acids Res. 12, 9383-9394, 1984  
 A: Title: Molecular cloning and cDNA sequence of human type III procollagen  
 A: Reference number: A93551; MUID:85087944; PMID:6096827  
 A: Accession: A93551  
 A: Molecule type: mRNA  
 A: Residues: 1065-1155, P', 1157-1466 <LOI>  
 A: Cross-references: EMBL:X01742; NID:G29584; PIDN:CAA5582.1  
 R: Miskin, M.; Dalgleish, R.; Kluge-Beckerman, B.; Reinhard, S.I.; Tolstoshev, P.; Brand  
 Biochemistry 25, 1408-1413, 1986  
 A: Title: Human type III collagen gene expression is coordinately modulated with the type  
 A: Reference number: 152393; MUID:86187804; PMID:3754462  
 A: Accession: 152393  
 A: Status: translated from GB/EMBL/DBJ  
 A: Molecule type: mRNA  
 A: Residues: 1161-1200 <MTS>  
 A: Cross-references: GB:M13146; NID:G180415; PIDN:AA52003.1; PID:G180416  
 R: Emanuel, B.S.; Cammizaro, L.A.; Seyer, U.M.; Myers, J.C.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985

A: Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long at  
 A: Reference number: 159025; MUID:85216505; PMID:3858826  
 A: Accession: 179359  
 A: Status: translated from GB/EMBL/DBJ  
 A: Molecule type: mRNA  
 A: Residues: 1165-1196 <EMA>  
 A: Cross-references: GB:M1134; NID:G180417; PIDN:AA52004.1; PID:G180418  
 R: Chu, M.L.; Weil, D.; de Wet, M.; Bernard, M.; Sippola, M.; Ramirez, F.  
 J. Biol. Chem. 260, 4357-4363, 1985  
 A: Title: Isolation of cDNA and genomic clones encoding human pro-alpha(III) collagen.  
 A: Reference number: A92516; MUID:85157600; PMID:2579949  
 A: Accession: A92516  
 A: Molecule type: DNA  
 A: Residues: 1176-1240, V', 1242-1356, P', 1358-1466 <CHU>  
 A: Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; C  
 A: Experimental source: liver  
 A: Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given  
 action  
 C: Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
 3-hydroxylated. About 1% of the lysines are 5-hydroxylated and some are subsequently  
 C: Genetics:  
 A: Gene: GDB:COL3A1  
 A: Cross-references: GDB:118729; OMIM:120180  
 A: Map position: 2q31-2q31  
 A: Introns: 27/1, 94/3, 111/3, 149/3, 176/3, 554/3, 587/3, 1175/3, 1275/1, 1337/3, 1418/  
 A: Note: the list of introns is incomplete; defects in this gene can result in Ehlers-D  
 C: Complex: type III collagen is a homotrimer of monomers initially linked by disulfide  
 er of their length, is formed with desmosine cross-links made from lysine and allysine  
 C: Function:  
 A: Description: structural component of extracellular fibrous polymer that maintains int  
 C: Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology  
 C: Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hy  
 F: 1-23/Domain: signal sequence #status predicted <EX>  
 F: 24-153/Domain: amino-terminal propeptide #status predicted <PRO>  
 F: 31-91/Domain: von Willebrand factor type C repeat homology <WVC>  
 F: 154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>  
 F: 154-167/Region: amino-terminal nonhelical telopeptide  
 F: 168-1196/Region: helical  
 F: 1091-1093/Region: cell attachment (R-G-D) motif  
 F: 1127-1221/Region: carboxyl-terminal nonhelical telopeptide  
 F: 1228-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>  
 F: 1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <RCC>  
 F: 24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicte  
 F: 153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted  
 F: 154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predict  
 F: 161,122/Modified site: allysine (Lys) #status predicted  
 F: 163,284,860,977,1106/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F: 263/Binding site: carboxylate (Lys) (covalent) #status experimental  
 F: 584,1094/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental  
 F: 948-949/Cleavage site: Gly-Ile (collagenase) #status experimental  
 F: 1106/Binding site: carboxylate (Lys) (covalent) #status predicted  
 Query Match 7.6%; Score 141.5; DB 1; Length 1466;  
 Best Local Similarity 27.6%; Pred. No. 0.042;  
 Matches 61; Conservative 13; Mismatches 74; Indels 73; Gaps 10;  
 QY 171 PGCAVPRKRVEMEMEDRCA-----AHSLEBQKXNLRGDPISNAT-----SRGAE 218  
 DB 844 PPGSGPAGPFPQGVKGRSGPGGAAGFGARGLPFGPSNPNPPPGSGSPGSDG 903  
 QY 219 PLPP-----PMPAVVAGAGGL-----ALLLVAGAGGMCWRERRA 256  
 DB 904 PPGPAGNTGAPSPGVSGPKDAGPGRKSGAGPAGPGLDAGTGAAGLAPPG 963  
 QY 257 KSESRRHP-----SFRGSGSLGIG--GGKMP-----REREPGLTAL 297  
 DB 964 MGPGRGSPGPQGVKESGPGANGLSGRRGPPGGLDGLAGTAGEPGRDGNPSDGLPG 1023  
 QY 298 RCG-----GAADPPCPHYEKSGDYCHPVLVQDGP 330  
 DB 1024 RDGSPGKGRDRENSPGAP-----GAPGHP-----GPP 1052

## RESULT 38

T32250  
 hypothetical protein T15B7.3 - *Caenorhabditis elegans*  
 C/Species: *Caenorhabditis elegans*  
 C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000  
 C/Accession: T32250  
 R/Pauley, A.; Gattung, S.  
 Submitted to the EMBL Data Library, September 1997  
 A/Description: The sequence of *C. elegans* cosmid T15B7.  
 A/Reference number: Z21139  
 A/Accession: T32250  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-319 <PAU>  
 A/Cross-references: EMBL:AF022985; PIDN:AB69959.1; GSPDB:GN00023; CESP:T15B7.3  
 A/Experimental source: strain Bristol N2; clone T15B7  
 C/Genetics:  
 A/Gene: CESP:T15B7.3  
 A/Map position: 5  
 A/Intons: 266/1  
 C/Superfamily: unassigned collagens

Query Match 7.6%; Score 141; DB 2; Length 319;  
 Best Local Similarity 25.6%; Pred. No. 0.0088;  
 Matches 90; Conservative 27; Mismatches 111; Indels 124; Gaps 23;

```

QY 13 VGALLLLGVGLVSGSL-----EPYVNSANKRPOAEG---YVLPQIGDRDLIC 62
DB 1 MSASTLVTAASAAGIAIVCVFTVMGFINDINSFYDEKIGELKEKFGYEQIA--WQAMI 58
QY 63 PRARPGPHSPNYEFYLYLVG-----GAQGRCEA-PPAPNLLITCDRDLRL 111
DB 59 PTPRSSGSS-----FLGRNKRQACNCGESRGCPCAPPP-----GQPG--- 101
QY 112 RFTIKQEVSPNMGHFRSHHDYIATSDGTREGLESIGVCLTRGMKYLRLVGSFP 171
DB 102 ---ARGAGLPIAGQ-----PGSGARINPARGRPFCIT-----CP 135
QY 172 RGGAVRKPESEMPERDRGAHSLPEKKNLPDPTSNATSGAAGPLPPSMRVAGA 231
DB 136 AGAPGPPGP-----PGA-----PGKGNNGQPGAPQS--GGRGP--PGRGP--ACD 177
QY 232 AGGIALLLGVAGAGGAMCWRRAKPSRSRHPG--PGSGRGG--SLGIGGGGG---MG 284
DB 178 AG-----SPGQGRHPSGNGRGGRGRGRGPGASGRGPGG 214
QY 285 PRARPGELGIALRGGAADP--PCPHYKVSQDYGHPIYTVODGPPGPPPN 335
DB 215 PAGA-PGQPG---RSGGAGTPGPGP---PGSGQPGHSGNDGVPTPGN 257

```

## RESULT 39

A34246  
 collagen alpha 1(VIII) chain precursor - rabbit  
 C/Species: *Oryctolagus cuniculus* (domestic rabbit)  
 C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C/Accession: A34246  
 R/Yamaguchi, N.; Benya, P. D.; van der Rest, M.; Nishimura, Y.  
 J. Biol. Chem. 264, 16022-16029, 1989  
 A/Title: The cloning and sequencing of alpha1(VIII) collagen cDNAs demonstrate that type  
 I remains similar to those of type X collagen.  
 A/Reference number: A34246; MID:89380199; PMID:2476437  
 A/Accession: A34246  
 A/Molecule type: mRNA  
 A/Residues: 1-744 <YAM>  
 A/Cross-references: GB:J05042; NID:G164895; PIDN:AAA1204.1; PID:G164895  
 C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
 F:1-20/Domain: signal sequence #status predicted <Sig>  
 F:21-744/Product: collagen alpha 1(VIII) chain #status predicted <MAT>  
 F:21-117/Region: amino-terminal nonhelical  
 F:118-571/Region: interrupted helical  
 F:572-744/Region: carboxyl-terminal nonhelical  
 F:617-743/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 7.6%; Score 141; DB 1; Length 744;  
 Best Local Similarity 23.6%; Pred. No. 0.022;  
 Matches 83; Conservative 29; Mismatches 108; Indels 132; Gaps 19;

```

QY 67 PPGHSSPNYEFYLYLVG-----GAQG-RCEAPPAANLLITCDRDLRLFT 114
DB 198 PPGHGLPG-----IGKPGGLPGQPGAKDRPKPPGP----- 233
QY 115 IKFQEVSPNMGHFRSHHDYIATSDGTR---EGLESIG--GYCLTRGMKYLRLVG 168
DB 224 ---PGLGPRK-----GEKGRMPGLPGLKGRPGKNGHGGPPGVLGCVG 272
QY 169 Q-----SPRGAVPRKPVSEMPERDRGAHSLPEKKNLPDPTSN 210
DB 273 KPQVTGFPQPGPLKPKGPPGPPGPGPIGVPGVQPGPGLPGVAKPGQDGPQPGFPG 332
QY 211 ATSRGAGPLPPEMPAIVA-----GAAGIALLLGVAGAGGAMCWRRR-----A 256
DB 333 KGEQGLPELPGPLPGVAKPGFPKPD---RGIQGVAGLGPGEKGPVGAPEGMG 387
QY 257 KPSRSRHPG--PGSGRGGSGL-----GGGGMGPR-----EAPGELGIALRG--GAAD 304
DB 388 PPGPGLPGIRGPMGPFGALGFPKXGEGIVGQGPFGKGEGLGPGKPGFLGTVG 447
QY 305 PP---PCPHYKVSQDYGH-----PYTVODGPP-----QSPNIT 336
DB 448 PPGIRGLPGPIGPKGEAGHKGLPGLPGVPLGLPKGEPGIPGPDGLGPGPI 499

```

## RESULT 40

S23298  
 collagen alpha 1(VIII) chain - chicken  
 C/Species: *Gallus gallus* (chicken)  
 C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C/Accession: S23298  
 R/Nishimura, Y.; Castagnola, P.; Gerecke, D.; Gordon, M.K.; Jacenko, O.; LuValle, P.; Mc  
 magh, N.; Olsen, B.R.  
 In Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp.79-114, Academic Pr.  
 A/Title: The molecular biology of collagens with short triple-helical domains.  
 A/Reference number: S22243  
 A/Accession: S23298  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-744 <NIN>  
 C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
 F:617-743/Domain: complement C1q carboxyl-terminal homology <C1Q>

Query Match 7.6%; Score 141; DB 1; Length 744;  
 Best Local Similarity 23.6%; Pred. No. 0.022;  
 Matches 83; Conservative 29; Mismatches 108; Indels 132; Gaps 19;

```

QY 67 PPGHSSPNYEFYLYLVG-----GAQG-RCEAPPAANLLITCDRDLRLFT 114
DB 198 PPGHGLPG-----IGKPGGLPGQPGAKDRPKPPGP----- 233
QY 115 IKFQEVSPNMGHFRSHHDYIATSDGTR---EGLESIG--GYCLTRGMKYLRLVG 168
DB 224 ---PGLGPRK-----GEKGRMPGLPGLKGRPGKNGHGGPPGVLGCVG 272
QY 169 Q-----SPRGAVPRKPVSEMPERDRGAHSLPEKKNLPDPTSN 210
DB 273 KPQVTGFPQPGPLKPKGPPGPPGPGPIGVPGVQPGPGLPGVAKPGQDGPQPGFPG 332
QY 211 ATSRGAGPLPPEMPAIVA-----GAAGIALLLGVAGAGGAMCWRRR-----A 256
DB 333 KGEQGLPELPGPLPGVAKPGFPKPD---RGIQGVAGLGPGEKGPVGAPEGMG 387
QY 257 KPSRSRHPG--PGSGRGGSGL-----GGGGMGPR-----EAPGELGIALRG--GAAD 304
DB 388 PPGPGLPGIRGPMGPFGALGFPKXGEGIVGQGPFGKGEGLGPGKPGFLGTVG 447
QY 305 PP---PCPHYKVSQDYGH-----PYTVODGPP-----QSPNIT 336

```

Db 448 PGIIRGLPGIPICPKGEAGHKGLPGVPGVGLGPKGSEPGIPGDOGLOQPGPI 499

## RESULT 41

hypothetical protein F26B1.4 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000

C/Accession: T30165

Riarteille, P.; Mamsley, P.; Kramer, J.

Submitted to the EMBL Data Library, November 1996

A/Description: The sequence of C. elegans cosmid F26B1.

A/Reference number: 220748

A/Accession: T30165

A/Status: Preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-305 <LAT>

A/Cross-references: EMBL:U08444; PIDN:AA337786.1; GSPDB:GN00019; CESP:F26B1.4

A/Experimental source: strain Bristol N2; clone F26B1

C/Genetics:

A/Gene: CESP:F26B1.4

A/Map position: 1

A/Introns: 51/3; 77/1

C/Superfamily: unassigned collagens

Query Match 7.6%; Score 140.5; DB 2; Length 305;

Best Local Similarity 25.8%; Pred. No. 0.0091;

Matches 69; Conservative 22; Mismatches 87; Indels 89; Gaps 15;

114 TIKFQEVSPNLMGHFRSHHDYIATSDG-----TRGLESLOGGVCLTRGMKVL 164

Db 47 TQHFQETSEKIV-----DLIVFATGETRBRKHNAITAPVPTVG----- 89

165 LRVGSGPGGAVPRKPVSEMERDGA-----AHSL-----EPKKNLPG----- 205

Db 90 FQTGTARNGCNCRCPPPPGPGNGEDGCRFYCNLSHLPLDPDGEFGVPSGSGDDIS 149

206 ----DP-----TSNATSRGAEGLPPSPMAVAAGAGLALLLVAGAGGAMCMRRRAK 257

Db 150 VLRHDPGCAQCPAPMPAPPPGRRPGVPGITGAPG-----LDGIPGHGTPGT----- 198

QY 258 PSESHHPR-GSFGRGSLGCGGGM-----GPREAPGELGIALRGGAADPP 307

Db 199 PCEPGLPGVBPDPGDKRPNNGNVVRIRIPRPPGRGAE-GDSGV---GG-----PG 250

QY 308 CPHYEKSGDYGHPIYITVQDGPQSP 334

Db 251 SPGLAGAPDGSAP-----GPGGP 270

## RESULT 42

hypothetical protein F52B1.4 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000

C/Accession: T22482

Riarteille, P.

Submitted to the EMBL Data Library, November 1996

A/Reference number: 219569

A/Accession: T22482

A/Status: Preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-304 <WIL>

A/Cross-references: EMBL:Z82268; PIDN:CAB05195.1; GSPDB:GN00022; CESP:F52B1.4

A/Experimental source: clone F52B1

C/Genetics:

A/Gene: CESP:F52B1.4

A/Map position: 4

A/Introns: 27/3

C/Superfamily: unassigned collagens

Query Match 7.6%; Score 140; DB 2; Length 304;

Best Local Similarity 26.8%; Pred. No. 0.0099;

Matches 59; Conservative 16; Mismatches 87; Indels 58; Gaps 10;

142 DGTREGLSTLOGVCL-----TRGMKVLIRVQSPRGAV-----P 177

Db 87 EGYVDQSEBAGQSGSCGCLPGAAGPAGTPG-----KGRPERPAAAGLPGRPPAP 141

QY 178 RRVSEPM-----ERDGAHSLRP-GKENLDPDTNATSGAAGPLP-----PSNP 226

Db 142 CEIRTPPCPCPCPGAPGAPGAPGPGDAGAPGAP-GQSGGAGAPGAPGKASGAPG 200

QY 227 AVAGANG--GLALLLVGAGAGAMCMRRRAKPESEHNP-----GPGSFGSGSISG 276

Db 201 GQAGAPGQADAGQSESTIGAPG-----QKPGQPGGAPGAPGPGGAPGAPG 253

QY 277 LGGGGMGPREAPPELIGALRGGAADPPPCPHYKVS 316

Db 254 PEGAPGQPADNPGAPGQAGAGGKICPKCAIDG 293

## RESULT 43

CGHUL

collagen alpha 1(X) chain precursor - human

N/Alternate names: procollagen alpha 1(X) chain

C/Species: Homo sapiens (man)

C/Date: 22-Nov-1993 #sequence\_revision 03-Nov-1995 #text\_change 22-Jun-1999

C/Accession: S26396; S30086; S18826; S18249; A43901; I51870; S21856

R/Reichenberger, E.; Beler, F.; Luvallie, P.; Olsen, B.R.; von der Mark, K.; Bertling, Y

FBS Lett. 311, 305-310, 1992

A/Title: Genomic organization and full-length cDNA sequence of human collagen X.

A/Reference number: S26396; MUID:93012005; PMID:1397333

A/Accession: S26396

A/Molecule type: DNA

A/Residues: 1-680 <RE1>

A/Cross-references: EMBL:X69952; EMBL:X72578; EMBL:X72579; EMBL:X72580; GB:S47714; GB:S

R/Apte, S.S.

Submitted to the EMBL Data Library, March 1992

A/Reference number: S30086

A/Accession: S30086

A/Molecule type: DNA

A/Residues: 717PGVCMVCVLL/52-680 <APT>

A/Cross-references: EMBL:X65120; NID:G23129

A/Note: the initial difference is probably due to translation of an intronic sequence

R/Apte, S.; Mattei, M.G.; Olsen, B.R.

FBS Lett. 282, 393-396, 1991

A/Title: Cloning of human alpha-1(X) collagen DNA and localization of the COL10A1 gene

A/Reference number: S15826; MUID:91243838; PMID:2037056

A/Accession: S15826

A/Molecule type: DNA

A/Residues: 561-647, 'G', 649-666 <AP2>

A/Cross-references: EMBL:X58879; NID:G30013; PIDN:CAA41686.1; PID:G30014

R/Thomas, J.T.; Creswell, C.J.; Rash, B.; Nicolai, H.; Jones, T.; Solomon, E.; Grant,

Biochem. J. 280, 617-623, 1991

A/Title: The human collagen X gene. Complete primary translated sequence and chromosome

A/Reference number: S18249; MUID:92109659; PMID:1764025

A/Accession: S18249

A/Molecule type: DNA

A/Residues: 1-26, 'T', 28-680 <THO>

A/Cross-references: EMBL:X60382; NID:G30094; PIDN:CAA42933.1; PID:G30095

A/Note: the sequence from Fig. 3 is inconsistent with that from Fig. 2 in having 532-A1

R/Reichenberger, E.; Aigner, T.; von der Mark, K.; Stoss, H.; Bertling, W.

Dev. Biol. 148, 562-572, 1991

A/Title: In situ hybridization studies on the expression of type X collagen in fetal hu

A/Reference number: A43901; MUID:9207285; PMID:1743401

A/Accession: A43901

A/Molecule type: mRNA

A/Residues: 547-656 <RE2>

A/Cross-references: GB:M74050; GB:D57494; NID:G339884; PIDN:AAA61221.1; PID:G553796

A/Note: sequence extracted from NCBI backbone (NCBI:69012, NCBI:69014)

R/Walls, G.A.; Rash, B.; Sweetman, W.A.; Thomas, J.T.; Super, M.; Evans, G.; Grant, M.

Am. J. Hum. Genet. 54, 169-178, 1994

A/Title: Amino acid substitutions of conserved residues in the carboxyl-terminal domain

pe Schmid.





A/Accession: S12851  
 A/Molecule type: mRNA  
 A/Residues: 1-911 <BRO>  
 R/Mayne, R.; Brewton, R.G.; Mayne, P.M.; Baker, J.R.  
 J. Biol. Chem. 268, 9381-9386, 1993  
 A/Title: Isolation and characterization of the chains of type V/type XI collagen present  
 A/Reference number: A4662; PMID:93252802; PMID:848632  
 A/Accession: C4662  
 A/Molecule type: protein  
 A/Residues: 347-354/356-600/912-924/925-961/962-998/999-1024 <MAX>  
 A/Experimental source: vitreous humor  
 A/Note: sequence modified after extraction from NCBI backbone  
 R/Mu, J.J.; Eyre, D.R.  
 J. Biol. Chem. 270, 18865-18870, 1995  
 A/Title: Structural analysis of cross-linking domains in cartilage type XI collagen. Ins  
 A/Reference number: A56978; PMID:95370194; PMID:7642541  
 A/Accession: A56978  
 A/Status: preliminary  
 A/Molecule type: protein  
 A/Residues: 206-229 <WDA>  
 R/Niyibizi, C.; Eyre, D.R.  
 Eur. J. Biochem. 224, 943-950, 1994  
 A/Title: Structural characteristics of cross-linking sites in type V collagen of bone.  
 A/Reference number: S48210; PMID:95010086; PMID:7925418  
 A/Accession: S65864  
 A/Molecule type: protein  
 A/Residues: 1-273-298 <NTX>  
 C/Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology  
 C/Keywords: hydroxyproline  
 F/211,223/Modified site: 4-hydroxyproline (Pro) #status experimental

Query Match 7.5%; Score 139; DB 2; Length 1024;  
 Best Local Similarity 24.3%; Pred. No. 0.043;  
 Matches 90; Conservative 19; Mismatches 118; Indels 144; Gaps 19;

```

QY 2 GPHSGPGGVAVGALLLLGLVGLVSGLSLEPYMNSANKRFQAGGYVLYPQIGDRLDL 61
DB 728 GPP--GPGGV-VQPGPFTBTG-----PIGR----- 751
QY 62 CPRARPPGHSSNNFFYLTVVGAQGRCAAPAPNLLTCRPPDLDFRTTFQBYIS 121
DB 752 -GHPGPPGPPGEGQ-----LPGAAGKGAQADPGQVSGKDP----- 789
QY 122 PNLGHEFRSHDYIATSDGREGLESLOGVCTRGKMLVVGOSPRGAVRKPV 181
DB 790 AGLRG-----PGR-GLPGAQAPGL-----KGGSPQGP 820
QY 182 SEWPERDRGAASL-----EPGKENLPDPTSNATSRGAEGLPPSPMPAVGAAG 234
DB 821 GPVSGPGEKSGAGTAGPIGLPGRPPGQPPG-PAGEKGAPEKGGQGPAGHDGVGPGV- 878
QY 235 LMLLLGVAGAGMCMRRRAKPSRSRHPG-----PGSFRG----- 272
DB 879 ----LPGRAPGAS-----PQEDGKGEIGERQKSGKDEGGQMGFGQDEBAG 925
QY 273 ----GSILGGGGKNGPPEAPFELGIALRGGAADPPFCFHYKVSADYGHPIYVOD 327
DB 926 PGPPIPIGLGPIG-PPEGEGNDVGPW----GPPGPP-GGIGIPGVEGQLPGAAGD 979
QY 328 GP--PGSPENI 336
DB 980 GPPGPIGPPGL 990

```

RESULT 47  
 A24450  
 collagen alpha 2 (VIII) chain - bovine (fragment)  
 N/Alternate names: 50K-A collagen (VIII), Desmect's membrane  
 C/Species: Bos pirtingensis taurus (cattle)  
 C/Date: 31-Mar-1988 #sequence \_revision 05-Apr-1995 #text\_change 03-May-1996  
 C/Accession: S12898; A24450  
 R/Mann, K.; Jander, R.; Korsching, E.; Kuehn, K.; Rautenberg, J.  
 FEBS Lett. 273, 168-172, 1990

A/Title: The primary structure of a triple-helical domain of collagen type VIII from bo  
 A/Reference number: S12898; PMID:91032171; PMID:2226849  
 A/Accession: S12898  
 A/Status: preliminary  
 A/Molecule type: protein  
 A/Residues: 1-469 <MAN>  
 R/Kapoor, R.; Bornstein, P.; Sage, E.H.  
 Biochemistry 25, 3930-3937, 1986  
 A/Title: Type VIII collagen from bovine Desmect's membrane: structural characterizat  
 A/Reference number: A90507; PMID:86296655; PMID:3527259  
 A/Accession: A24450  
 A/Molecule type: protein  
 A/Residues: 1-24 <KAP>  
 C/Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

Query Match 7.5%; Score 138.5; DB 2; Length 469;  
 Best Local Similarity 25.3%; Pred. No. 0.02;  
 Matches 69; Conservative 17; Mismatches 72; Indels 115; Gaps 14;

```

QY 147 GLESLG---GVCITRGKVLIRVQSPRGAVP-----RKPSEMPERDRGA 193
DB 216 GLPGQPPGAXGEPGRGPPGLI---GPTGYMGGLPGXGDRPAGVPLGDRG-- 269
QY 194 HSLRPGKENTPGD-----PTSNAT-----SRGAEGLPPSPMPAVGAAG 223
DB 270 ---EPGDEGPGQPGQGLGPPGLPSAGLPGRRAVPGXGERTGPGVFGIRDDG 326
QY 234 -GLALLLVAGAGAGMCMRRRAKPSRSRHPG-----GSPGSG 272
DB 327 PSLGA---GKPLPG-----ERGLPGAHPGPGTGPXGPEFTGRPGGPAAGALGX 376
QY 273 GSILG-----GGGGMGR-----EAPRGLILRG 300
DB 377 GDLPGQPLRPPSGIPGIQGPAGPIPGGLPLGXGPGLPPEPGEKVGEPVA--- 432
QY 301 GAADPPCPHYEKVSGDPHYPIYVODGPPSP 333
DB 433 GPPGPPGVPSPGLTGPBPFP-----GPPGP 459

```

RESULT 48  
 A41182  
 collagen alpha 1(II) chain precursor - mouse  
 C/Species: Mus musculus (house mouse)  
 C/Date: 28-May-1992 #sequence \_revision 28-May-1992 #text\_change 13-Aug-1999  
 C/Accession: A41182; A44885  
 R/Metsieranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.  
 J. Biol. Chem. 266, 16862-16869, 1991  
 A/Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and  
 A/Reference number: A41182; PMID:91358489; PMID:1885613  
 A/Accession: A41182  
 A/Status: preliminary; not compared with conceptual translation  
 A/Molecule type: DNA  
 A/Residues: 1-1419 <MCT>  
 A/Cross-references: GB:M65161  
 R/Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.  
 Development 111, 945-953, 1991  
 A/Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to cartil  
 A/Reference number: A44885; PMID:91347939; PMID:1879363  
 A/Accession: A44885  
 A/Molecule type: DNA  
 A/Residues: 1-28 <CHE>  
 A/Cross-references: GB:S63190; MID:9234368; PIDN:AAB19627.1; PID:9234369  
 A/Note: sequence extracted from NCBI backbone (NCBIN:63190, NCBI:P:63192)  
 C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology  
 C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trit  
 F/1191-1419/domain: fibrillar collagen carboxyl-terminal homology <FOC>

Query Match 7.5%; Score 138.5; DB 2; Length 1419;  
 Best Local Similarity 30.1%; Pred. No. 0.066;  
 Matches 58; Conservative 13; Mismatches 75; Indels 47; Gaps 9;

QY 173 GAVPRKPYSEWPERDRGAASLEPGKENTL-----PGD--PTSNATRGAEGLPPSP 224



```

Db 430 GGAAPFPPGGERGAPGRNGF-----PGQDGLAGPKAPGSRGSGLAGPPGAGNDGPRPQ 484
QY 225 MPVAVAGAGGATALLLLGVAGAGAMCWRRRRAPSRSRHPGP-----GSFGRGSLGLGCG 280
Db 485 EPLGPGARG-----LTCRPGDAGPQGVPSGAPGEDGRGPGPQDQARGQCPVWGFPPG 539
QY 281 GGMGPRRAEPGEGLGIA-----LRG-----GGAADPPFCPHYKVSQDYGHPIYIVQDGP 329
Db 540 KGANGEPGKKGKGLAAGPGLRGLPGKDGRTGAAGPPGPGPAGERGEGQAP-----GP 593
QY 330 -----PQSP 334
Db 594 SGFGGLGPPGPP 606

RESULT 49
B41182
collagen alpha 1(II) chain precursor (long splice form) - mouse
C/Species: Mus musculus (house mouse)
C/Date: 28-May-1992 #sequence_revision 28-May-1992 #ext_change 16-Jul-1993
C/Accession: B41182
R/Metasequent, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
J. Biol. Chem. 266, 16862-16869, 1991
A/Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
A/Reference number: A41182; PMID:91358489; PMID:1885613
A/Accession: B41182
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: DNA
A/Residues: 1-1487 <MET>
A/Cross-references: GB:M65161
C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C/Keywords: alternative splicing; coll; extracellular matrix; glycoprotein; trine
F:33-91/Domain: von Willebrand factor type C repeat homology <YMC>
F:1259-1487/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 7.5%; Score 138.5; DB 2; Length 1487;
Best Local Similarity 30.1%; Pred. No. 0.069;
Matches 58; Conservative 13; Mismatches 75; Indels 47; Gaps 9;

QY 173 GGAAPRKEVEMEMERDGAHSLERKENTL-----PGD--PTSMATRGAEGLPPPS 224
Db 498 GGAAPFPPGGERGAPGRNGF-----PGQDGLAGPKAPGSRGSGLAGPPGAGNDGPRPQ 552
QY 225 MPVAVAGAGGATALLLLGVAGAGAMCWRRRRAPSRSRHPGP-----GSFGRGSLGLGCG 280
Db 553 EPLGPGARG-----LTCRPGDAGPQGVPSGAPGEDGRGPGPQDQARGQCPVWGFPPG 607
QY 281 GGMGPRRAEPGEGLGIA-----LRG-----GGAADPPFCPHYKVSQDYGHPIYIVQDGP 329
Db 608 KGANGEPGKKGKGLAAGPGLRGLPGKDGRTGAAGPPGPGPAGERGEGQAP-----GP 661
QY 330 -----PQSP 334
Db 662 SGFGGLGPPGPP 674

RESULT 50
CGH02E
collagen alpha 2(XI) chain precursor - human (fragment)
N/Alternate names: procollagen alpha 2(XI) chain
N/contains: proline/arginine-rich protein (PARP)
C/Species: Homo sapiens (man)
C/Date: 07-Jun-1990 #sequence_revision 03-Oct-1995 #ext_change 22-Jun-1999
C/Accession: S34790; A32645
R/Zhidkova, N.I.; Brewton, R.G.; Mayne, R.
FEBS Lett. 326, 25-28, 1993
A/Title: Molecular cloning of PARP (proline/arginine-rich protein) from human cartilage
chain.
A/Reference number: S34790; PMID:93314796; PMID:8325574
A/Accession: S34790
A/Molecule type: mRNA
A/Residues: 1-663 <ZHI>

```

```

A/Cross-references: EMBL:L19987; NID:9306439; PIDN:AAA35498.1; PID:9306440
R.Kimura, T.; Cheah, K.S.E.; Chan, S.D.H.; Lui, V.C.H.; Mattei, M.G.; van der Rest, M.;
J. Biol. Chem. 264, 13910-13916, 1989
A/Title: The human alpha2(XI) collagen (COL11A2) chain. Molecular cloning of cDNA and g
A/Reference number: A32645; PMID:93314796; PMID:12760050
A/Accession: A32645
A/Molecule type: DNA, mRNA
A/Residues: 586-1546 <KIM>
A/Cross-references: GB:U04974; NID:9180714; PIDN:AAA52034.1; PID:9180715
A/Note: parts of this sequence were determined by protein sequencing
C/Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C/Genetics:

```

```

A/Gene: GDB:COL11A2
A/Cross-references: GDB:119788; OMIM:120290
A/Map position: 6p21.3-6p21.3
A/Introns: 1302/3; 1320/3; 1332/3; 1350/3; 1440/1; 1477/3
A/Note: the list of introns is incomplete
C/Complex: type XI collagen may be a heterotrimer of two alpha 1(XI) chains (see PIR:CG
3(XI) chain (see PIR:CGH02E), initially linked by disulfide bonds among their carboxyl
med with desmosome cross-links made from lysine and allysine residues
C/Function:
A/Description: structural component of extracellular fibrous polymer associated with ce
A/Note: may play a role in controlling the lateral growth of collagen II fibrils
C/Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
C/Keywords: coll; extracellular matrix; glycoprotein; hydroxylysine; hydroxyprol
F:1-354/Domain: non-collagenous (fragment) #status predicted <NC3>
F:1-187/Product: proline/arginine-rich PKP protein (fragment) #status predicted <PARP>
F:1255-305/Domain: collagenous, triple helix #status predicted <COL2>
F:306-342/Domain: non-collagenous #status predicted <NC2>
F:343-1356/Region: helical
F:429-441/Region: cell attachment (R-G-D) motif
F:447-449/Region: cell attachment (R-G-D) motif
F:1257-1259/Region: cell attachment (R-G-D) motif
F:1357-1380/Region: carboxyl-terminal nonpeptide (fragment) #status predicted <CPT>
F:1381-1546/Domain: carboxyl-terminal nonpeptide (fragment) #status predicted <CPT>
F:1403-1546/Domain: fibrillar collagen carboxyl-terminal homology (fragment) #status at
F:109-163; 1511-1545/disulfide bonds: #status predicted
F:319/Modified site: allysine (lys) #status predicted
F:432; 1266/Modified site: 5-hydroxylysine (lys) #status predicted
F:432; 1266/Binding site: carboxylate (lys) (covalent) #status predicted
F:927; 933; 1008; 1017; 1035; 1038; 1290; 1295; 1305; 1317; 1320/Modified site: 4-hydroxyproline
F:928/Modified site: 4-hydroxyproline (Pro) #status atypical
F:942; 1023; 1299/Modified site: 5-hydroxylysine (lys) #status experimental
F:942; 1023; 1299/Binding site: carboxylate (lys) (covalent) #status experimental
F:1427; 1433; 1450; 1459/disulfide bonds: interchain #status predicted
F:1460/Binding site: carboxylate (Asn) (covalent) #status predicted

```

```

Query Match 7.5%; Score 138.5; DB 1; Length 1546;
Best Local Similarity 31.7%; Pred. No. 0.072;
Matches 64; Conservative 18; Mismatches 71; Indels 49; Gaps 14;

QY 147 GLBSLOGSVCLTRGMKYLRLVQSGSPRGAVPRK--PYSEMEMERDRGAHSLERKENTLP 204
Db 1183 GSEGRQGG---KAK-----GDPGALGAPGKTPVGP-----AGPAGKGPDPGLR 1224
QY 205 GDPTS-----NATSR-GAEGPLPPSPMPVAVAGAGGALL-----LLGVAGAGAMCWR 252
Db 1225 GLPGSVGQGRPGATNGAGPPGPGPLPGLRBDAGAKSGKGGHPIGLIGPGG----- 1279
QY 223 RRRAPKSSSRHPG--GFRGRGSLGL--CGGGCMGPRRAEPGEGLGIA--LRGGAGADPPFCP 309
Db 1280 -EQGKGDGRGLPGQSGGQGGEMGIPGAGAPGP--GAPPLPAPAPKAKGATGPGGP 1357
QY 310 HYER-VSGDYGHPIYIVQDGP 330
Db 1338 KGEKGVGPPEHP-----GPP 1353

```

Search completed: February 11, 2003, 12:06:08  
Job time : 36.805 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2003, 12:05:10 / Search time 14.5409 Seconds  
(without alignments)  
597.393 Million cell updates/sec

Title: US-10-021-121-4

Perfect score: 1850  
Sequence: 1 MGPPIGSGPGVGVGALLILG.....PYIVQDGPFGSPNITYKV 340

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 140259 seqs, 25548876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 50 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1850	100.0	340	12	US-10-021-121-4
2	1846	99.8	340	9	US-10-138-787-3
3	1841	99.5	455	12	US-10-021-121-2
4	632	34.2	346	12	US-10-021-121-9
5	629.5	34.0	333	10	US-09-754-105-2
6	629.5	34.0	333	10	US-09-978-339-2
7	629.5	34.0	333	12	US-10-021-121-10
8	622.5	33.6	333	9	US-10-138-787-4
9	608.5	32.9	345	9	US-10-138-787-5
10	498	26.9	89	9	US-10-138-787-13
11	498	26.9	89	10	US-09-862-179A-17
12	284.5	15.4	92	10	US-09-864-761-48262
13	205.5	11.1	136	10	US-09-864-761-48257
14	202.5	10.9	106	10	US-09-925-297-639
15	201	10.9	82	9	US-10-138-787-11
16	201	10.9	82	9	US-09-862-179A-15
17	200.5	10.8	82	9	US-10-138-787-12
18	200.5	10.8	82	10	US-09-862-179A-16
19	179	9.7	238	10	US-09-904-954-2

20	176	9.5	209	9	US-09-921-984-2	Sequence 2, Appli
21	174.5	9.4	233	9	US-10-138-787-7	Sequence 7, Appli
22	172.5	9.3	218	10	US-09-925-297-510	Sequence 510, App
23	169.5	9.2	201	10	US-09-904-954-4	Sequence 4, Appli
24	169	9.1	201	9	US-10-138-787-8	Sequence 8, Appli
25	168.5	9.1	209	9	US-10-138-787-6	Sequence 6, Appli
26	167.5	9.1	228	8	US-08-578-684-4	Sequence 4, Appli
27	166	9.0	205	9	US-10-138-787-10	Sequence 10, Appli
28	164.5	8.9	204	9	US-10-001-054-48	Sequence 48, Appli
29	164.5	8.9	204	9	US-10-028-072-288	Sequence 288, App
30	164.5	8.9	204	9	US-10-121-049-288	Sequence 288, App
31	164.5	8.9	204	9	US-10-123-904-288	Sequence 288, App
32	164.5	8.9	204	9	US-10-140-470-288	Sequence 288, App
33	164.5	8.9	204	9	US-10-175-746-288	Sequence 288, App
34	164.5	8.9	204	9	US-10-176-918-288	Sequence 288, App
35	164.5	8.9	204	9	US-10-176-921-288	Sequence 288, App
36	160.5	8.7	228	9	US-10-138-787-9	Sequence 9, Appli
37	154	8.3	27	10	US-09-925-301-1225	Sequence 1225, Ap
38	141	7.6	28	10	US-09-862-179A-2	Sequence 2, Appli
39	138	7.5	638	9	US-10-001-867-108	Sequence 108, App
40	130.5	7.1	380	10	US-09-823-240-10	Sequence 10, Appli
41	130.5	7.1	674	10	US-09-925-299-979	Sequence 979, App
42	129	7.0	595	9	US-09-854-133-187	Sequence 187, App
43	129	7.0	595	10	US-09-738-973-187	Sequence 187, App
44	129	7.0	1806	10	US-09-919-497-56	Sequence 56, Appli
45	126.5	6.8	403	10	US-09-925-302-689	Sequence 689, App
46	126.5	6.8	520	9	US-09-978-295A-614	Sequence 614, App
47	126.5	6.8	520	9	US-09-978-697-614	Sequence 614, App
48	126.5	6.8	520	9	US-09-978-192A-614	Sequence 614, App
49	126.5	6.8	520	9	US-09-999-832A-614	Sequence 614, App
50	126.5	6.8	520	9	US-09-978-189-614	Sequence 614, App

## ALIGNMENTS

RESULT 1  
US-10-021-121-4  
Sequence 4, Application US/10021121  
Patient No. US2002014244A1  
GENERAL INFORMATION:  
APPLICANT: Caras, Ingrid W  
TITLE OF INVENTION: A2-1 Neurotrophic Factor  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 MB floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Winpatin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/021,121  
FILING DATE: 06-Dec-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/635,130  
FILING DATE: 19-Mar-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Torchia, PhD., Timothy E.  
REGISTRATION NUMBER: 36,700  
REFERENCE/DOCKET NUMBER: P1001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/952-9881  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 340 amino acids

TYPE: Amino Acid  
TOPOLOGY: Linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 4  
US-10-021-121-4

Query Match 100.0%; Score 1850; DB 12; Length 340;  
Best Local Similarity 100.0%; Pred. No. 8,7e-126;  
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPPHSGPGVAVGALLLIGVGLVSGLSLEPYWNSANKRFOAEGGYLYPQIGDRIDL 60  
DB 1 MGPPHSGPGVAVGALLLIGVGLVSGLSLEPYWNSANKRFOAEGGYLYPQIGDRIDL 60  
QY 61 LCPRAAPPGHSSPNVEFYKLYLVGAQGRCEAPAPNLLTCRPPDLRFTTKFOEY 120  
DB 61 LCPRAAPPGHSSPNVEFYKLYLVGAQGRCEAPAPNLLTCRPPDLRFTTKFOEY 120  
QY 121 SPNLWGEHFRSHHDYIIATSDGTREGLESLOGGVCLTRGMKYLRLVQSGPRGAVPRKP 180  
DB 121 SPNLWGEHFRSHHDYIIATSDGTREGLESLOGGVCLTRGMKYLRLVQSGPRGAVPRKP 180  
QY 181 VSEMPERDRGAHSLPEKKNLPDPTSNATSRGAEGPLPPSMPPAVAGAAGLALLL 240  
DB 181 VSEMPERDRGAHSLPEKKNLPDPTSNATSRGAEGPLPPSMPPAVAGAAGLALLL 240  
QY 241 GVAAGAGACWRRRAKPSERSRHPGSGFRGSGSLGLGGGGMGPREAPGELIALRG 300  
DB 241 GVAAGAGACWRRRAKPSERSRHPGSGFRGSGSLGLGGGGMGPREAPGELIALRG 300  
QY 301 GAADPPCPHYEKVSGDYGHPIYIVODGPPSPNIIYKYV 340  
DB 301 GAADPPCPHYEKVSGDYGHPIYIVODGPPSPNIIYKYV 340

RESULT 2  
US-10-138-787-3

; Sequence 3, Application US/10138787  
; Patent No. US20020172984A1  
; GENERAL INFORMATION:  
; APPLICANT: Holland, Sacha  
; APPLICANT: Mbamalu, Geraldine  
; APPLICANT: Pawsom, Tony  
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED  
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR  
; FILE REFERENCE: 11757.23USWO  
; CURRENT APPLICATION NUMBER: US/10/138,787  
; CURRENT FILING DATE: 2002-05-03  
; PRIOR APPLICATION NUMBER: US/09/214,631  
; PRIOR FILING DATE: 1999-03-12  
; PRIOR APPLICATION NUMBER: PCT/CA97/00473  
; PRIOR FILING DATE: 1997-07-04  
; PRIOR APPLICATION NUMBER: 60/021,272  
; PRIOR FILING DATE: 1996-07-05  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 3  
; LENGTH: 340  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-138-787-3

Query Match 99.8%; Score 1846; DB 9; Length 340;  
Best Local Similarity 99.7%; Pred. No. 1,7e-125;  
Matches 339; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPPHSGPGVAVGALLLIGVGLVSGLSLEPYWNSANKRFOAEGGYLYPQIGDRIDL 60  
DB 1 MGPPHSGPGVAVGALLLIGVGLVSGLSLEPYWNSANKRFOAEGGYLYPQIGDRIDL 60  
QY 61 LCPRAAPPGHSSPNVEFYKLYLVGAQGRCEAPAPNLLTCRPPDLRFTTKFOEY 120  
DB 61 LCPRAAPPGHSSPNVEFYKLYLVGAQGRCEAPAPNLLTCRPPDLRFTTKFOEY 120

QY 121 SPNLWGEHFRSHHDYIIATSDGTREGLESLOGGVCLTRGMKYLRLVQSGPRGAVPRKP 180  
DB 121 SPNLWGEHFRSHHDYIIATSDGTREGLESLOGGVCLTRGMKYLRLVQSGPRGAVPRKP 180  
QY 181 VSEMPERDRGAHSLPEKKNLPDPTSNATSRGAEGPLPPSMPPAVAGAAGLALLL 240  
DB 181 VSEMPERDRGAHSLPEKKNLPDPTSNATSRGAEGPLPPSMPPAVAGAAGLALLL 240  
QY 241 GVAAGAGACWRRRAKPSERSRHPGSGFRGSGSLGLGGGGMGPREAPGELIALRG 300  
DB 241 GVAAGAGACWRRRAKPSERSRHPGSGFRGSGSLGLGGGGMGPREAPGELIALRG 300  
QY 301 GAADPPCPHYEKVSGDYGHPIYIVODGPPSPNIIYKYV 340  
DB 301 GAADPPCPHYEKVSGDYGHPIYIVODGPPSPNIIYKYV 340

RESULT 3

US-10-021-121-2  
; Sequence 2, Application US/10021121  
; Patent No. US2002014244A1  
; GENERAL INFORMATION:  
; APPLICANT: Caras, Ingrid W  
; TITLE OF INVENTION: A2-1 Neurotrophic Factor  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Minipatin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/021,121  
; FILING DATE: 06-Dec-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/635,130  
; FILING DATE: 19-Mar-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Torchia, PhD., Timothy E.  
; REGISTRATION NUMBER: 36,700  
; REFERENCE/DOCKET NUMBER: P1001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-8674  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 455 amino acids  
; TYPE: Amino Acid  
; TOPOLOGY: Linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-10-021-121-2

Query Match 99.5%; Score 1841; DB 12; Length 455;  
Best Local Similarity 100.0%; Pred. No. 5,3e-125;  
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPPHSGPGVAVGALLLIGVGLVSGLSLEPYWNSANKRFOAEGGYLYPQIGDRIDL 60  
DB 1 MGPPHSGPGVAVGALLLIGVGLVSGLSLEPYWNSANKRFOAEGGYLYPQIGDRIDL 60  
QY 61 LCPRAAPPGHSSPNVEFYKLYLVGAQGRCEAPAPNLLTCRPPDLRFTTKFOEY 120  
DB 61 LCPRAAPPGHSSPNVEFYKLYLVGAQGRCEAPAPNLLTCRPPDLRFTTKFOEY 120  
QY 121 SPNLWGEHFRSHHDYIIATSDGTREGLESLOGGVCLTRGMKYLRLVQSGPRGAVPRKP 180

```

Db 121 SPNMGHFRSHHDYIIATSDGTREGLESLGGVCLTRGKVLRAVCGSPRGAVPRKX 180
Qy 181 VSEMPMRDGAHSLRPGKENLPDPTSNATSRGAGCPLPPSPMPVAGAAGLALLLL 240
Db 181 VSEMPMRDGAHSLRPGKENLPDPTSNATSRGAGCPLPPSPMPVAGAAGLALLLL 240
Qy 241 GVAGAGAMCRRRRRAKPSRHPGSPFGAGSLGAGGGGGMGPRAEPGELGALRG 300
Db 241 GVAGAGAMCRRRRRAKPSRHPGSPFGAGSLGAGGGGGMGPRAEPGELGALRG 300
Qy 301 GAADPPCPHYEKVSGDYGHPVYIVDGPPOSPPNITY 338
Db 301 GAADPPCPHYEKVSGDYGHPVYIVDGPPOSPPNITY 338

```

## RESULT 4

```

US-10-021-121-9
; Sequence 9, Application US/10021121
; Patent No. US2002014244A1
; GENERAL INFORMATION:
; APPLICANT: Caras, Ingrid M
; TITLE OF INVENTION: A2-1 Neurotrophic Factor
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPacIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/021,121
; FILING DATE: 06-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/635,130
; FILING DATE: 19-Mar-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Torchia, Phd., Timothy E.
; REGISTRATION NUMBER: 36,700
; REFERENCE/DOCKET NUMBER: P1001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/952-9861
; TELEFAX: 650/952-9861
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 346 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-10-021-121-9

```

```

Query Match 34.2%; Score 632; DB 12; Length 346;
Best Local Similarity 39.5%; Pred. No. 2,1e-38;
Matches 145; Conservative 48; Mismatches 116; Indels 58; Gaps 9;
Qy 8 PGGRVAGALLLVGLVSGV-----SLBPVYVNSANKRFOAGGVLYVPOIGRDLRL 61
Db 4 PGGRVAGALLLVGLVSGV-----SLBPVYVNSANKRFOAGGVLYVPOIGRDLRL 61
Qy 62 CPRARPPGSSRYEYKYLVVGAAGRCAPAPANLLTCDRDLRFTIKFOEYS 121
Db 64 CPRARPPGSSRYEYKYLVVGAAGRCAPAPANLLTCDRDLRFTIKFOEYS 118
Qy 122 PNLGHEFRSHHDYIIATSDGTREGLESLGGVCLTRGKVLRAVCGSPRGAVPRKXV 181
Db 119 PNLGHEFRSHHDYIIATSDGTREGLESLGGVCLTRGKVLRAVCGSPRGAVPRKXV 178

```

```

Qy 182 SEMPERDGAHSLR-PGKENLPDPTSNATSRGAGCPLPPSPMPVAGAAGL----- 236
Db 179 SEMPERDGAHSLR-PGKENLPDPTSNATSRGAGCPLPPSPMPVAGAAGL----- 231
Qy 237 -----LVLGVAAGGA-----MCMRRRAKPSRHPGSPFGAGSLG 277
Db 232 GFNSKVALFAVAGAGVIFLLIIIFLVLLKLRKRRTQ-----RAALSL 282
Qy 278 -----GGGGMGPRAEPGELGALRGGAADPPCPHYEKVSGDYGHPVYIVDGPPOS 333
Db 283 STLASPKGSSGAGTSPSDIILPLR--TTNNYCPHYEKVSGDYGHPVYIVDGPPOS 339
Qy 334 PNIYKYV 340
Db 340 ANIYKYV 346

```

## RESULT 5

```

US-09-754-105-2
; Sequence 2, Application US/09754105
; Patent No. US20010009768A1
; GENERAL INFORMATION:
; APPLICANT: Cerretti, Douglas
; APPLICANT: Reddy, Pranthia
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING CYTOKINE DESIGNATED LERK-5
; FILE REFERENCE: 28232
; CURRENT FILING DATE: 2001-01-03
; PRIOR APPLICATION NUMBER: 09/329,531
; PRIOR FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent version 3.0
; SEQ ID NO 2
; LENGTH: 333
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-754-105-2

```

```

Query Match 34.0%; Score 629.5; DB 10; Length 333;
Best Local Similarity 40.9%; Pred. No. 3,1e-38;
Matches 135; Conservative 52; Mismatches 130; Indels 13; Gaps 5;
Qy 14 GALLLVGLVSGVSLBPVYVNSANKRFOAGGVLYVPOIGRDLRLCPRARPPGSS 73
Db 14 GALLLVGLVSGVSLBPVYVNSANKRFOAGGVLYVPOIGRDLRLCPRARPPGSS 70
Qy 74 PNYEYKYLVVGAAGRCAPAPANLLTCDRDLRFTIKFOEYSPNLGHFRSH 133
Db 71 GQYEVYKYVMDKQADRCITIKENTPLNCAKPDQDIKFTIKFOEYSPNLMGLEFQNK 130
Qy 134 DYIIATSDGTREGLESLGGVCLTRGKVLRAVCGSPRGAVPRKXVSEMER-DR 190
Db 131 DYIIATSDGTREGLESLGGVCLTRGKVLRAVCGSPRGAVPRKXVSEMER-DR 190
Qy 191 GAHSLRPGKENLPDPTSNATSRGAGCPLPPSPMPVAGAAGLALLLVGAAGAMC 250
Db 191 GRSSTSPFVKNPPOSSSTIDNSAGHSNNILGSEVALFAGIASCIIFIVITLVVLL 250
Qy 251 WRRRAKPSRHPGSPFGAGSLGAGGGGGMGPRAEPGELGALRGGAADPPCPH 310
Db 251 KYRRHRHGPSQHTTSLSLTLATPKRSNN--NGSEPSDIIILP--TAUSVCFH 303
Qy 311 YEKVSGDYGHPVYIVDGPPOSPPNITYKV 340
Db 304 YEKVSGDYGHPVYIVDGPPOSPPNITYKV 333

```

## RESULT 6

```

US-09-978-339-2
; Sequence 2, Application US/09978339
; Patent No. US20020103358A1
; GENERAL INFORMATION:

```

```

APPLICANT: Cerretti, Douglas P.
Inventor: Raddy, Pranthiba
TITLE OF INVENTION: No. US20020103358A1el Cyclokinine Designated Lerk-5
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESSES:
ADDRESSER: Immunex Corporation
STREET: 51 University Street
City: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98101

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple 7.1
SOFTWARE: Microsoft Word, Version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/978,339
FILING DATE: 15-Oct-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/271,948
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Seese, Kathryn A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2823
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
TELEX: 756822

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 333 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-978-339-2

Query Match 34.0%; Score 629.5; DB 10; Length 333;
Best Local Similarity 40.9%; Pred. No. 3,1e-38;
Matches 135; Conservative 52; Mismatches 110; Indels 13; Gaps 5

QY 14 GALLLLGVGLVSGGLELPEYVWNSANKRFOAGGVTLYPQIGRLDLLCPARPQGHSS 73
Db 14 GVLVALLCRLTAISKSIVLEPIYWNSSNSKELPQGLVLYPQIGDLIDLICPKV--DSKTV 70
QY 74 PNYEFYLYLVGAQGRCEAPPAEPLLLTCRPPDLIRFTKFOEYSPVLMGHEPSSH 133
Db 71 GQHYHYKYNVWDQADRCITIKKENTPLNCKAKPDQDIKFTIKQESPVLMGLEPQYK 130
QY 134 DYVIATSDGTREGLESLLQGVVCLTRGMKVLRLVQ--SPRGAVPKRPVSEMPER-DR 190
Db 131 DYVIATSDGTREGLESLLQGVVCLTRGMKVLRLVQ--SPRGAVPKRPVSEMPER-DR 190
QY 191 GAASISLPGKEMLPDPTNATNSRGAEGELPPSPMAYAGAAGLALLLGVAAGAGANC 250
Db 191 GRSSITSPYKPNPGSGSTGNSGSHGNNILGSEVALFAGIASCIITFIYITLLVLL 250
QY 251 WRERRAPSPSRPAGSGSPGRGSLGLGGGGMGPRAEPGLGIALRGGAADPFPCF 310
Db 251 KYRRRRHKHSRPHITTLISLTLATPRSGNN-----NGSEPDITIIPLR---TDSVFCF 303
QY 311 YEKYSGPYGHPIYITVODGPPQSPNITYYV 340
Db 304 YEKVSGDYGHPIYITVQEMPPQSPNITYYV 333

```

```

GENERAL INFORMATION:
  APPLICANT: Caras, Ingrid M
  TITLE OF INVENTION: A2-1 Neurotrophic Factor
  NUMBER OF SEQUENCES: 10
  CORRESPONDENCE ADDRESS:
    ADDRESSER: Genentech, Inc.
    STREET: 1 DNA Way
    CITY: South San Francisco
    STATE: California
    COUNTRY: USA
    ZIP: 94080

COMPUTER READABLE FORM:
  MEDIUM TYPE: 3.5 inch, 1.44 MB floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: WinPatIn (Genentech)

CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/10/021,121
  FILING DATE: 06-Dec-2001
  CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US/08/635,130
  FILING DATE: 19-Mar-1996

ATTORNEY/AGENT INFORMATION:
  NAME: Torchia, PhD., Timothy E.
  REGISTRATION NUMBER: 36,700
  REFERENCE/DOCKET NUMBER: P1001
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 650/225-8674
    TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 10:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 333 amino acids
    TYPE: Amino Acid
    TOPOLOGY: Linear

SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-021-121-10

Query Match      34.0%  Score 629.5,  DB 12,  Length 333,
Best Local Similarity 40.9%,  Pred. No. 3,1e-38;
Matches 135;  Conservative 52;  Mismatches 130;  Indels 13;  Gaps 5,

QY  14  GALLILGVIGVSGSLSEPVVWNSNKRFEQLEGVIVYPOIGRLDILCPRAPRPHSS 73
Db  14  GVLIMVLCRRALIKSIVLEPIYVNSSNSKFLPGGLVIVYPOIGRLDILCKV--DSKIV 70
QY  74  PNYEFKLYIVGAGQGRRCRAPPAENILALTCRBDLDLRFITKOEYSPLWGHERRSH 133
Db  71  GQVEYKVMVMDQADNRRTIKENTPLINCXKPDQDIKFTIKQBSPLWGLGEQKK 130
QY  134  DYYIITSGTREGLESLOGGVCLTRGKVLRLRGQ--SRGGAVPKPVSEMPMR-DR 190
Db  131  DYYIITSGTSGSEGLDNEGVCQFRRAKILMKVGQDASSAGSTRNDPFRPELAGTN 190
QY  191  GAHSLSEPGKEMLPQDPFNSNATSRGAEBPLPPMPMAVAGGLALLLGVAGAGAMC 250
Db  191  GRSTISPFYKKNPSSSTDGNSAGHSNNILIGSEVALFAGIASGCIIFYIITIVLL 250
QY  251  WRRRAKPEESRHPQBSGFRGSGSLGLGGGGMGPRLAEFGELGIALRGGAADPPFCR 310
Db  251  KYRRRRRKHSPOHTTTLSTLATPFRKSGNN---NGSEPSDIIIPUR--TADSVFCR 303
QY  311  YEKVSGDYGHPIYIVODGPOSPPIITYKV 340
Db  304  YEKVSGDYGHPIYIVODGPOSPPIITYKV 333

RESULT 8
US-10-138-787-4
; Sequence 4, Application US/10138787
; Patent No. US20020172984A1
GENERAL INFORMATION:
  APPLICANT: Holland, Sacha

```

RESULT 9  
 US-10-138-787-5  
 ; Sequence 5, Application US/10138787  
 ; Patent No. US20020172984A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Holland, Sacha  
 ; APPLICANT: Mbemalu, Geraldine  
 ; APPLICANT: Pawson, Tony  
 ; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED  
 ; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ETK-RELATED RECEPTOR  
 ; TITLE OF INVENTION: TYROSINE KINASES  
 ; FILE REFERENCE: 11757.23USMO  
 ; CURRENT APPLICATION NUMBER: US/10/138,787  
 ; CURRENT FILING DATE: 2002-05-03  
 ; PRIOR APPLICATION NUMBER: US/09/214,631  
 ; PRIOR FILING DATE: 1999-03-12  
 ; PRIOR APPLICATION NUMBER: PCT/CA97/00473  
 ; PRIOR FILING DATE: 1997-07-04  
 ; PRIOR APPLICATION NUMBER: 60/021,272  
 ; PRIOR FILING DATE: 1996-07-05  
 ; NUMBER OF SEQ ID NOS: 13  
 ; SOFTWARE: PatentIn Ver. 2.0

Query Match	26.9%;	Score 498;	DB 9;	Length 69;
Best Local Similarity	100.0%;	Pred. No. 1.9e-29;		
Matches 89;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	252	RRRAAKPSSESHPPPGSGFGRGSGTGLGGGGGMRREAPPELIGTALNGGGAAADPPPCPHY	311	

Db 1 RRRRAKPESSRRPBGSGSGSLGIGGGGNGPREAPGELGIALRGGAADPPCPHY 60  
QY 312 EKVSDDYGHPIYIVODGPPSPNNIYKYV 340  
Db 61 EKVSDDYGHPIYIVODGPPSPNNIYKYV 89

RESULT 11  
US-09-862-179A-17  
Sequence 17, Application US/09862179A  
Patent No. US20020147306A1  
GENERAL INFORMATION:  
APPLICANT: Lin, Danny  
APPLICANT: Pawson, Anthony  
TITLE OF INVENTION: PEPTIDES THAT MODULATE THE INTERACTION OF B CLASS EPHRINS  
FILE REFERENCE: MTSI-P01-009  
CURRENT APPLICATION NUMBER: US/09/862,179A  
CURRENT FILING DATE: 2001-05-21  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 17  
LENGTH: 89  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-862-179A-17

Query Match 26.9%; Score 498; DB 10; Length 89;  
Best Local Similarity 100.0%; Pred. No. 1.9e-29;  
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 RRRRAKPESSRRPBGSGSGSLGIGGGGNGPREAPGELGIALRGGAADPPCPHY 311  
Db 1 RRRRAKPESSRRPBGSGSGSLGIGGGGNGPREAPGELGIALRGGAADPPCPHY 60

QY 312 EKVSDDYGHPIYIVODGPPSPNNIYKYV 340  
Db 61 EKVSDDYGHPIYIVODGPPSPNNIYKYV 89

RESULT 12  
US-09-864-761-48262  
Sequence 48262, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecm1ca-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 48262  
LENGTH: 92  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL136092.7  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9  
OTHER INFORMATION: SWISSPROT HIT: P98172, EVALU 4.00e-51  
OTHER INFORMATION: EST\_HUMAN HIT: BE562822.1, EVALU 3.00e-50  
US-09-864-761-48262

Query Match 15.4%; Score 284.5; DB 10; Length 92;  
Best Local Similarity 54.2%; Pred. No. 4.1e-14;  
Matches 52; Conservative 15; Mismatches 24; Indels 5; Gaps 1;

QY 42 FQAEAGYVLYPOIGRDLDCPRAPPGHSSPNYEFYLYVGAQGRCEAPPAPNLL 101  
Db 1 FLSGKLVLYPKIGDLIDICRAEAGR-----YEYLYLVREQAAGCTVADPNLL 55

QY 102 LTCDPDLDRFTIKQOYSPNLMGHEFRSHDYTI 137  
Db 56 VTCNRPEQIRFTIKQOESPNYMGLEFKGHDIYI 91

RESULT 13  
US-09-864-761-48257  
Sequence 48257, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecm1ca-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/609,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 48257
; LENGTH: 136
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION: MAP TO A1136092.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: EST_HUMAN HIT: AA377505.1, EVALUATE 2.00e-26
; OTHER INFORMATION: SWISSPROT HIT: P52795, EVALUATE 4.00e-40
US-09-864-761-48257

Query Match      11.1%; Score 205.5; DB 10; Length 136;
Best Local Similarity 37.7%; Pred. No. 3e-08; Indels 39; Gaps 5;
Matches 52; Conservative 13; Mismatches 34;

QY 230 GAAGGAA-----LTLGAGAGGA-----MCMRRRRAKPSRHRPQ 266
DB 11 GAGGGSGDDPDGPFNGKVALFAVAGACVIFLLIIFLTVLLKTKRKRRKHTQQ----- 65
QY 267 GSGRGSGSLG-----GGGGMGPREAPPELGIALRGGAADPPFPFHETKXSGYGHV 322
DB 66 ----RAAALSLSTLASPKGSGTAGTEPSDIIPLR---TTNNYCPHYEKVSGYGHV 118
QY 323 YIVDGPQSPFNITYKV 340
DB 119 YIVQEMPQSPFNITYKV 136

RESULT 14
US-09-925-297-639
; Sequence 639, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 639
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-297-639
```

```

Query Match      10.9%; Score 202.5; DB 10; Length 106;
Best Local Similarity 43.6%; Pred. No. 3.7e-08; Indels 25; Gaps 4;
Matches 48; Conservative 11; Mismatches 26;

QY 235 LALLLVAGAGGACWRRRRRAKPSRHRPQSGFGSLG-----GGGGMGPREAP 230
DB 18 LTVLLKTK-----RKRRKHTQQ-----RAAALSLSTLASPKGSGTAGTEP 59
QY 291 GELGIALRGGAADPPFPFHETKXSGYGHVYIVDGPQSPFNITYKV 340
DB 60 SDIIPLR---TTNNYCPHYEKVSGYGHVYIVQEMPQSPFNITYKV 106

RESULT 15
US-10-138-787-11
; Sequence 11, Application US/10138787
; Patent No. US20020172984A1
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Madanlu, Geraldine
; APPLICANT: Pawsen, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/10/138,787
; CURRENT FILING DATE: 2002-05-03
; PRIOR APPLICATION NUMBER: US/09/214,631
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: PCT/CA97/00473
; PRIOR FILING DATE: 1997-07-04
; PRIOR APPLICATION NUMBER: 60/021,272
; PRIOR FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-138-787-11

Query Match      10.9%; Score 201; DB 9; Length 82;
Best Local Similarity 47.3%; Pred. No. 3.6e-08; Indels 16; Gaps 3;
Matches 44; Conservative 9; Mismatches 24;

QY 252 RRRRAKPSRHRPQSGFGSLG-----GGGGMGPREAPPELGIALRGGAADPP 307
DB 2 RRRRKHQ-----RAAALSLSTLASPKGSGTAGTEPSDIIPLR---TTNNY 49
QY 308 CPHYEKVSGYGHVYIVDGPQSPFNITYKV 340
DB 50 CPHYEKVSGYGHVYIVQEMPQSPFNITYKV 82

RESULT 16
US-09-862-179A-15
; Sequence 15, Application US/09862179A
; Patent No. US20020147306A1
; GENERAL INFORMATION:
; APPLICANT: Lin, Danny
; APPLICANT: Pawsen, Anthony
; TITLE OF INVENTION: PEPTIDES THAT MODULATE THE INTERACTION OF B CLASS EPHRINS
; TITLE OF INVENTION: AND PZ DOMAINS
; FILE REFERENCE: MTSI-P01-009
; CURRENT APPLICATION NUMBER: US/09/862,179A
; CURRENT FILING DATE: 2001-05-21
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 15
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-862-179A-15
```



US-09-862-179A-15

10.9%; Score 201; DB 10; Length 82;

Query Match

Best Local Similarity 47.3%; Pred. No. 3,6e-08;

Matches 44; Conservative 9; Mismatches 24; Indels 16; Gaps 3;

QY 252 RRRAKPSESRRPFGSGFSGSLG---GGGGMGPREAPGELGIALRGGAADPPF 307

DB 2 RRRHRRHTQC-----RAALSLSTLASKGSGTAGTEPSDIIPLR---TTENNY 49

QY 308 CPYKVGSDYGHPIYIVODGPPSPNNIYKYV 340

DB 50 CPYKVGSDYGHPIYIVODGPPSPNNIYKYV 82

RESULT 17

US-10-138-787-12

Sequence 12, Application US/10138787

Patent No. US20020172984A1

GENERAL INFORMATION:

APPLICANT: Holland, Sacha

APPLICANT: Mhamadi, Geraldine

APPLICANT: Pawson, Tony

TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED

BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR

TITLE OF INVENTION: TYROSINE KINASES

FILE REFERENCE: 11757.23USWO

CURRENT APPLICATION NUMBER: US/10/138,787

CURRENT FILING DATE: 2002-05-03

PRIOR APPLICATION NUMBER: US/09/214,631

PRIOR FILING DATE: 1999-03-12

PRIOR APPLICATION NUMBER: PCT/CA97/00473

PRIOR FILING DATE: 1997-07-04

PRIOR APPLICATION NUMBER: 60/021,272

PRIOR FILING DATE: 1996-07-05

NUMBER OF SEQ ID NOS: 13

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO: 12

LENGTH: 82

TYPE: PRT

ORGANISM: Homo sapiens

US-10-138-787-12

Query Match

Best Local Similarity 10.8%; Score 200.5; DB 9; Length 82;

Matches 38; Conservative 4; Mismatches 8; Indels 3; Gaps 1;

QY 288 AEPGELGIALRGGAADPPCPHYKVGSDYGHPIYIVODGPPSPNNIYKYV 340

DB 33 SEPSDIIPLR---TADSVFCHYKVGSDYGHPIYIVODGPPSPNNIYKYV 82

RESULT 18

US-09-862-179A-16

Sequence 16, Application US/09862179A

Patent No. US20020147306A1

GENERAL INFORMATION:

APPLICANT: Lin, Danny

APPLICANT: Pawson, Anthony

TITLE OF INVENTION: PEPTIDES THAT MODULATE THE INTERACTION OF B CLASS EPHRINS

FILE REFERENCE: MTST-P01-009

CURRENT APPLICATION NUMBER: US/09/862,179A

CURRENT FILING DATE: 2001-05-21

NUMBER OF SEQ ID NOS: 44

SOFTWARE: PatentIn version 3.1

SEQ ID NO: 16

LENGTH: 82

TYPE: PRT

ORGANISM: Homo sapiens

US-09-862-179A-16

Query Match

10.8%; Score 200.5; DB 10; Length 82;

Best Local Similarity 71.7%; Pred. No. 3,9e-08;

Matches 38; Conservative 4; Mismatches 8; Indels 3; Gaps 1;

QY 288 AEPGELGIALRGGAADPPCPHYKVGSDYGHPIYIVODGPPSPNNIYKYV 340

DB 33 SEPSDIIPLR---TADSVFCHYKVGSDYGHPIYIVODGPPSPNNIYKYV 82

RESULT 19

US-09-904-954-2

Sequence 2, Application US/09904954

Patent No. US20020010325A1

GENERAL INFORMATION:

APPLICANT: BECKMANN, M. P.

TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE

RECEPTOR HEK

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: IMMUNEX CORPORATION

STREET: 51 UNIVERSITY STREET

CITY: SEATTLE

STATE: WASHINGTON

COUNTRY: USA

ZIP: 98101

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: Apple Macintosh

OPERATING SYSTEM: Apple System 7.1

SOFTWARE: Microsoft Word for Apple, Version 5.1a

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/904,954

FILING DATE: 12-Jul-2001

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/240,124

FILING DATE: &lt;Unknown&gt;

APPLICATION NUMBER: US 08/114,426

FILING DATE: 30-AUG-1993

APPLICATION NUMBER: US 08/109,745

FILING DATE: 20-AUG-1993

ATTORNEY/AGENT INFORMATION:

NAME: SEESE, KATHRYN A.

REGISTRATION NUMBER: 32,172

REFERENCE/DOCKET NUMBER: 2814-C

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 587-0430

TELEFAX: (206) 233-0644

TELEX: 756822

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 238 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-904-954-2

Query Match

Best Local Similarity 9.7%; Score 179; DB 10; Length 238;

Matches 65; Conservative 28.4%; Pred. No. 4,4e-06; Mismatches 80; Indels 60; Gaps 12;

QY 7 GPGGVRVALLLIGVIGLSLEPYVNSANKRQAGGYLYLQIGDRDLDP--R 64

DB 24 GPG-----GALG-----NRHAVVNSNQHLLRR--GYTVGVVNDYDLYCPHYN 67

QY 65 ARPPGPHSP---NYEYLYLVGAQGRCEAPAPNILLTCDRPDL---DLRFTIKY 117

DB 68 SSGVPGAGPFGGGAEGYLYLVVNSNGYRTCAAGGPK--RMEGRPHAPHSIKSEKF 126

QY 118 QEYSPNIMGEHPSHHDYIATSDGTREGLESLOGGVCULTRGMKYLKRVGSPRGANV 177

DB 127 QRYSAFLGYEFHAGHEVYIISTPTNLH-----WKCLR--MKVFVCCASTSHSG--- 174

QY 178 RKPVSMP-----MERDRGAHSL-----PGKENLP 204  
 Db 175 EKPVPITLPQFTMGPNKINVLDEPFGENPQVPLKLSISGTSKREHLP 223

## RESULT 20

US-09-921-984-2  
 Sequence 2, Application US/09921984  
 Patent No. US20020156239A1

## GENERAL INFORMATION:

APPLICANT: Flanagan, John G.

Cheng, Hwai-Jong

TITLE OF INVENTION: Eph Receptor Ligands, and Uses Related  
 There to

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD

STREET: 60 State Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCII(txt)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/921,984

FILING DATE: 31-Jul-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/308,814

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Vincent, Matthew P.

REGISTRATION NUMBER: 36,709

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 227-7400

TELEFAX: (617) 227-5941

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 209 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULAR TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-921-984-2

Query Match 9.5%; Score 176; DB 9; Length 209;

Best Local Similarity 29.3%; Pred. No. 6.2e-06;

Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;

QY 33 VYVNSANKRFAE-----GGYVLYPQIGRLDLICPRARPGRHSPNVEFYLYLVGA 87

Db 35 VYVNSRNPQVSAVGGGYVEVSINDYLDICPHYGAFLP-PABRMERYLLVYVNGE 93

QY 88 QGRRCAPPAVNLITCDRDL--DIRFTIKQVSPNLMGHEPFSHEDYIIAASDGT 144

Db 94 GHASCDHRQGFKEKWEKRNPAAPGGLKFSKQLFTFSLGFEFRGHEYYISATP-- 151

QY 145 REGLESLOGVCLTRGKVLRLVQSPRGAVPRKPVSEMPMERDRGAHSLPGKENLP 204

Db 152 -----PNLVDRPCLR--LKYVYR-----PTMETLY 174

QY 205 GDP---TSNATSRGAE 218

Db 175 EAPPIFTSNSSCSGLG 192

RESULT 21

## US-10-138-787-7

Sequence 7, Application US/10138787

Patent No. US2002017294A1

GENERAL INFORMATION:

APPLICANT: Holland, Sacha

APPLICANT: Madamaly, Geraldine

APPLICANT: Pawsen, Tony

TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED

BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR

TITLE OF INVENTION: TYROSINE KINASES

FILE REFERENCE: 11757.21USWO

CURRENT APPLICATION NUMBER: US/10/138,787

CURRENT FILING DATE: 2002-05-03

PRIOR APPLICATION NUMBER: US/09/214,631

PRIOR FILING DATE: 1999-03-12

PRIOR APPLICATION NUMBER: PCT/CA97/00473

PRIOR FILING DATE: 1997-07-04

PRIOR APPLICATION NUMBER: 60/021,272

PRIOR FILING DATE: 1996-07-05

NUMBER OF SEQ ID NOS: 13

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 7

LENGTH: 233

TYPE: PRT

ORGANISM: Homo sapiens

US-10-138-787-7

Query Match 9.4%; Score 174.5; DB 9; Length 233;

Best Local Similarity 27.4%; Pred. No. 9e-06;

Matches 60; Conservative 24; Mismatches 74; Indels 61; Gaps 11;

QY 20 GVLGIVSLSELPYVNSANKRFAEGGYLYPQIGRLDLICR-----ARPGHSPN 75

Db 27 GALTG-----NRHAYVWSSNOHLRE-GTYQVNVNDYLDICHTNSSGAGPBGGAZ 80

QY 76 YEFYLYLVG-----GAQGR--CEAPPAVNLITCDRDLIRFTIKQVSPNLMG 127

Db 81 YLVVWSANGYRTCAASGQFRWECNRPNAPH-----SPKFSKPRYSAFSLGY 131

QY 128 EPRSHEDYIITATSDGTREGLESLOGVCLTRGKVLRLVQSPRGAVPRKPVSEMP-- 185

Db 132 EFHAGHEYYIISTPHNLH-----WKCLR--MKVFCASSTHSG--EKPVPITLP 179

QY 186 -----MERDRGAHSL-----PGKENLP 204

Db 180 TMGPVVKINVLDEPGENPQVPLKLSISGTSKREHLP 218

RESULT 22

US-09-925-297-510

Sequence 510, Application US/09925297

Patent No. US20020081659A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA105

CURRENT APPLICATION NUMBER: US/09/925,297

CURRENT FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: PCT/US00/05989

PRIOR FILING DATE: 2000-03-08

PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1999-03-12

NUMBER OF SEQ ID NOS: 928

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 510

LENGTH: 218

TYPE: PRT

ORGANISM: Homo sapiens

US-09-925-297-510

Query Match 9.3%; Score 172.5; DB 10; Length 218;

Best Local Similarity 28.3%; Pred. No. 1.2e-05;

Matches 51; Conservative 30; Mismatches 80; Indels 19; Gaps 6;

```

QY 8 PGQVVGALLLGVLTGSL---EPVYNSANKRFQAGGVLYLPQIGDRLLDLCPR 64
DB 8 PGQAVAMEFLMPLIGLCCSLAADRHTVFWSSNPKFRNE-DYTHQVLDYDILCPH 66
QY 65 APPGPHSPN--YFFKLYLVGAQGRCEAPPAFNLLTCRDL---DLRTIKQE 119
DB 67 YE---DHSVADAMQYILLYVEHEEYQUCQFQSDQVRCQNRSAHGFEKLSERQR 123
QY 120 YSPNMGHFRSHHDYIATSDGTREGESLQGVCLTRGMKYLRLVQSPRGAVPRK 179
DB 124 FTPTLGKEFKGSHSYYSKPIHQHEDR-----CRLKATVSGKITHTSPQAHNDNGE 176

RESULT 23
US-09-904-954-4
/ Sequence 4, Application US/09904954
/ Patent No. US20020010325A1
/ GENERAL INFORMATION:
/ APPLICANT: BECKMAN, M. P.
/ CERRETTI, DOUGLAS P.
/ TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE
/ RECEPTOR HEAT
/ NUMBER OF SEQUENCES: 4
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: TMMNEX CORPORATION
/ STREET: 51 UNIVERSITY STREET
/ CITY: SEATTLE
/ STATE: WASHINGTON
/ COUNTRY: USA
/ ZIP: 98101
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: Apple Macintosh
/ OPERATING SYSTEM: Apple System 7.1
/ SOFTWARE: Microsoft Word for Apple, Version 5.1a
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/904,954
/ FILING DATE: 12-Jul-2001
/ CLASSIFICATION: <unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/240,124
/ FILING DATE: <unknown>
/ APPLICATION NUMBER: US 08/114,426
/ FILING DATE: 30-AUG-1993
/ APPLICATION NUMBER: US 08/109,745
/ FILING DATE: 20-AUG-1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: SEESE, KATHRYN A.
/ REGISTRATION NUMBER: 32,172
/ REFERENCE/DOCKET NUMBER: 2814-C
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 587-0430
/ TELEFAX: (206) 233-0644
/ TELEX: 756822
/ INFORMATION FOR SEQ ID NO: 4:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 201 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-904-954-4

Query Match 9.2%; Score 169.5; DB 10; Length 201;
Best Local Similarity 29.9%; Pred. No. 1.7e-05;
Matches 66; Conservative 18; Mismatches 82; Indels 55; Gaps 10;

QY 24 LVSGSL-EPVYNSANKRFQAGGVLYLPQIGDRLLDLCPRAPPGHSGSPNIEFYKLY 82
DB 20 LRGSSLRHVYVWNSNPRL-LRGDAVVELGNDYLDIVCPHYEGPPEGP--ETFALY 76
QY 83 LVGAQGRCEAP-PAFNLLTCRDLRTIKQEYSPLMGHFRSHHDYIATSD 141

```

```

DB 77 MYMVPYEQAGLGPFAVYRWVCSLPFGHVQFSEKIQRTFTPSLGEFFPGSTYYISVP 136
QY 142 DGTREGESLQGVCLTRGMKYLRLVQSPRGAVPRKXVSEMPERDRGAHSLPEPKE 201
DB 137 --TPE-----SSGQCL-----RLOVSCCKERKESAHVP----- 164
QY 202 NLRGDPSTNATS--RGAGPLPPSPMPAVAGAAGLALLL 240
DB 165 ---GSPGEGTSGMRGDTTSP-----LCILL 189

RESULT 24
US-10-138-787-8
/ Sequence 8, Application US/10138787
/ Patent No. US20020172984A1
/ GENERAL INFORMATION:
/ APPLICANT: Holland, Sacha
/ APPLICANT: Mbamalu, Geraldine
/ APPLICANT: Pawsen, Tony
/ TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
/ TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
/ FILE REFERENCE: 11757.23USMO
/ CURRENT APPLICATION NUMBER: US/10/138,787
/ CURRENT FILING DATE: 2002-05-03
/ PRIOR APPLICATION NUMBER: US/09/214,631
/ PRIOR FILING DATE: 1999-03-12
/ PRIOR APPLICATION NUMBER: PCT/CA97/00473
/ PRIOR FILING DATE: 1997-07-04
/ PRIOR APPLICATION NUMBER: 60/021,272
/ PRIOR FILING DATE: 1996-07-05
/ NUMBER OF SEQ ID NOS: 13
/ SOFTWARE: Patentln Ver. 2.0
/ SEQ ID NO 8
/ LENGTH: 201
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-138-787-8

Query Match 9.1%; Score 169; DB 9; Length 201;
Best Local Similarity 29.8%; Pred. No. 1.9e-05;
Matches 70; Conservative 19; Mismatches 84; Indels 62; Gaps 11;

QY 17 LILGVY-----GLVSGSL-EPVYNSANKRFQAGGVLYLPQIGDRLLDLCPRAPP 68
DB 6 LILTVMAAFISPLRGSSLRHVYVWNSNPR-SLRGDAVVELGNDYLDIVCPHYEGP 64
QY 69 GPRSSNVEFYKLYLVGAQGRCEAP-PAFNLLTCRDLRTIKQEYSPLMGH 127
DB 65 GPPEGP--ETFALYVDMWGYEQAGLGPGRYRWVCSLPFGHVQFSEKIQRTFTPSLGF 122
QY 128 EFRSHHDYIATSDGTREGESLQGVCLTRGMKYLRLVQSPRGAVPRKXVSEMPYE 187
DB 123 EFLPGERTYYISVP--TPE-----SSGQCL-----RLOVSCCKE 155
QY 188 RDRGAHSLPEPKENLPDPTSNATS--RGAGPLPPSPMPAVAGAAGLALLL 240
DB 156 RKESAHVP-----GSPGEGTSGMRGDTTSP-----LCILL 189

RESULT 25
US-10-138-787-6
/ Sequence 6, Application US/10138787
/ Patent No. US20020172984A1
/ GENERAL INFORMATION:
/ APPLICANT: Holland, Sacha
/ APPLICANT: Mbamalu, Geraldine
/ APPLICANT: Pawsen, Tony
/ TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
/ TITLE OF INVENTION: BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
/ FILE REFERENCE: 11757.23USMO

```



APPLICANT: Baker, Kevin  
APPLICANT: Goddard, Audrey  
APPLICANT: Gurney, Austin  
APPLICANT: Hebert, Carolyn  
APPLICANT: Henzel, William  
APPLICANT: Kabakoff, Rhona  
APPLICANT: Shelton, David  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin  
APPLICANT: Wood, William  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC  
FILE REFERENCE: P3034R1PCT  
CURRENT APPLICATION NUMBER: US/10/001,054  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/079689  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079920  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/082999  
PRIOR FILING DATE: 1998-04-24  
PRIOR APPLICATION NUMBER: 60/083545  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/085149  
PRIOR FILING DATE: 1998-05-12  
PRIOR APPLICATION NUMBER: 60/087607  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/088858  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/090691  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/096891  
PRIOR FILING DATE: 1998-08-17  
PRIOR APPLICATION NUMBER: 60/096894  
PRIOR FILING DATE: 1998-08-17  
PRIOR APPLICATION NUMBER: 60/099803  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/100263  
PRIOR FILING DATE: 1998-09-14  
PRIOR APPLICATION NUMBER: 60/100390  
PRIOR FILING DATE: 1998-09-15  
PRIOR APPLICATION NUMBER: 60/101476  
PRIOR FILING DATE: 1998-09-23  
PRIOR APPLICATION NUMBER: 60/107783  
PRIOR FILING DATE: 1998-11-10  
PRIOR APPLICATION NUMBER: 60/108649  
PRIOR FILING DATE: 1998-11-18  
PRIOR APPLICATION NUMBER: 60/112420  
PRIOR FILING DATE: 1998-12-15  
PRIOR APPLICATION NUMBER: 60/113296  
PRIOR FILING DATE: 1998-12-22  
PRIOR APPLICATION NUMBER: 60/115554  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/115558  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/116533  
PRIOR FILING DATE: 1999-01-20  
PRIOR APPLICATION NUMBER: 60/123618  
PRIOR FILING DATE: 1999-03-10  
PRIOR APPLICATION NUMBER: 60/131294  
PRIOR FILING DATE: 1999-04-07  
PRIOR APPLICATION NUMBER: 60/140650  
PRIOR FILING DATE: 1999-06-22  
PRIOR APPLICATION NUMBER: 60/141037  
PRIOR FILING DATE: 1999-06-23  
PRIOR APPLICATION NUMBER: 60/144758  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: 60/162506  
PRIOR FILING DATE: 1999-10-29  
PRIOR APPLICATION NUMBER: 60/170262  
PRIOR FILING DATE: 1999-12-09

PRIOR APPLICATION NUMBER: 60/187202  
PRIOR FILING DATE: 2000-03-03  
PRIOR APPLICATION NUMBER: 60/209832  
PRIOR FILING DATE: 2000-06-05  
PRIOR APPLICATION NUMBER: 60/232887  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: 09/180997  
PRIOR FILING DATE: 1998-11-19  
PRIOR APPLICATION NUMBER: 09/218517  
PRIOR FILING DATE: 1998-12-22  
PRIOR APPLICATION NUMBER: 09/284291  
PRIOR FILING DATE: 1999-04-12  
PRIOR APPLICATION NUMBER: 09/380137  
PRIOR FILING DATE: 1999-08-25  
PRIOR APPLICATION NUMBER: 09/380138  
PRIOR FILING DATE: 1999-08-25  
PRIOR APPLICATION NUMBER: 09/380913  
PRIOR FILING DATE: 1999-09-09  
PRIOR APPLICATION NUMBER: 09/403297  
PRIOR FILING DATE: 1999-10-18  
PRIOR APPLICATION NUMBER: 09/423741  
PRIOR FILING DATE: 1999-11-10  
PRIOR APPLICATION NUMBER: 09/709238  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 09/802706  
PRIOR FILING DATE: 2001-03-09  
PRIOR APPLICATION NUMBER: 09/866034  
PRIOR FILING DATE: 2001-05-25  
PRIOR APPLICATION NUMBER: 09/872035  
PRIOR FILING DATE: 2001-06-01  
PRIOR APPLICATION NUMBER: 09/882636  
PRIOR FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: 09/918585  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: 09/924419  
PRIOR FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: 09/927796  
PRIOR FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: 09/929404  
PRIOR FILING DATE: 2001-08-13  
PRIOR APPLICATION NUMBER: 09/941992  
PRIOR FILING DATE: 2001-08-25  
PRIOR APPLICATION NUMBER: 09/946374  
PRIOR FILING DATE: 2001-09-04  
PRIOR APPLICATION NUMBER: PCT/US98/18824  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: PCT/US99/00106  
PRIOR FILING DATE: 1999-01-05  
PRIOR APPLICATION NUMBER: PCT/US99/05028  
PRIOR FILING DATE: 1999-03-08  
PRIOR APPLICATION NUMBER: PCT/US99/08615  
PRIOR FILING DATE: 1999-04-20  
PRIOR APPLICATION NUMBER: PCT/US99/12252  
PRIOR FILING DATE: 1999-06-02  
PRIOR APPLICATION NUMBER: PCT/US99/20111  
PRIOR FILING DATE: 1999-09-01  
PRIOR APPLICATION NUMBER: PCT/US99/20594  
PRIOR FILING DATE: 1999-09-08  
PRIOR APPLICATION NUMBER: PCT/US99/28313  
PRIOR FILING DATE: 1999-11-30  
PRIOR APPLICATION NUMBER: PCT/US99/28551  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/28634  
PRIOR FILING DATE: 1999-12-01  
PRIOR APPLICATION NUMBER: PCT/US99/30095  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: PCT/US99/30999  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US00/00376  
PRIOR FILING DATE: 2000-01-06  
PRIOR APPLICATION NUMBER: PCT/US00/03565  
PRIOR FILING DATE: 2000-02-11  
PRIOR APPLICATION NUMBER: PCT/US00/04341

```

; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: PCT/US00/04342
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: PCT/US00/06884
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: PCT/US00/13705
; PRIOR FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: PCT/US00/14941
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: PCT/US00/15264
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: PCT/US00/22031
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: PCT/US00/23522
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: PCT/US00/30873
; PRIOR FILING DATE: 2000-11-10
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: PCT/US01/06666
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: PCT/US01/17092
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: PCT/US01/17800
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US01/19692
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: PCT/US01/21066
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: PCT/US01/21735
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: PCT/US01/27099
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 91
; SEQ ID NO 48
; TYPE: PRT
; LENGTH: 204
; ORGANISM: Homo Sapien
US-10-001-054-48

Query Match      8.9%; Score 164.5; DB 9; Length 204;
Best Local Similarity 27.7%; Pred. No. 4,1e-05;
Matches 46; Conservative 31; Mismatches 74; Indels 15; Gaps 5;

QY 18 LIGVLGLVGLSLLEPYRMSANKRFQAEQGYLYTPOIGRLDLCPARPAPPHSS-PHY 76
DB 8 LIGLCCSLAADHNTYFMNSNKNPNE-DYTHQVLANVYVILICPHYE--DHADADM 63
QY 77 ERYKLYLVGAQGRCEAPAPNLLITCDPDL--DLRFITKPEYSNMLNGHEFRSHH 133
DB 64 EGYILVLVHEEQYQLCQPSKQYVRWCNRPSPAKHGPEKLSEKFGFTFTLGEKEFGH 123
QY 134 DYIITATSDGTREGLSLGGVGLTRGMKYLRLVYGSPRGCAVPRK 179
DB 124 SYIYISKPIHQHEDR-----CLRKLTVYSGKITSPAHNPDQ 162

RESULT 29
US-10-028-072-288
; Sequence 288, Application US/10028072
; Publication No. US20030004311A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
```

```

; APPLICANT: Denoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Geriltsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Guirney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Matanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang
; TITLE OF INVENTION:
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/028,072
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059836
; PRIOR FILING DATE: 1997-09-24
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062285
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062287
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062814
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/062816
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063045
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063082
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/063127
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063327
; PRIOR FILING DATE: 1997-10-27
; PRIOR APPLICATION NUMBER: 60/063329
; PRIOR FILING DATE: 1997-10-27
; PRIOR APPLICATION NUMBER: 60/063550
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063561
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063704
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063733
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063735
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063738
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063755
```

PRIOR FILING DATE:	1997-10-17
PRIOR APPLICATION NUMBER:	60/064248
PRIOR FILING DATE:	1997-11-03
PRIOR APPLICATION NUMBER:	60/064809
PRIOR FILING DATE:	1997-11-07
PRIOR APPLICATION NUMBER:	60/065186
PRIOR FILING DATE:	1997-11-12
PRIOR APPLICATION NUMBER:	60/065846
PRIOR FILING DATE:	1997-11-17
PRIOR APPLICATION NUMBER:	60/066364
PRIOR FILING DATE:	1997-11-21
PRIOR APPLICATION NUMBER:	60/066453
PRIOR FILING DATE:	1997-11-24
PRIOR APPLICATION NUMBER:	60/066511
PRIOR FILING DATE:	1997-11-24
PRIOR APPLICATION NUMBER:	60/066770
PRIOR FILING DATE:	1997-11-24
PRIOR APPLICATION NUMBER:	60/069212
PRIOR FILING DATE:	1997-12-11
PRIOR APPLICATION NUMBER:	60/069278
PRIOR FILING DATE:	1997-12-11
PRIOR APPLICATION NUMBER:	60/069334
PRIOR FILING DATE:	1997-12-11
PRIOR APPLICATION NUMBER:	60/069594
PRIOR FILING DATE:	1997-12-16
PRIOR APPLICATION NUMBER:	60/072320
PRIOR FILING DATE:	1998-01-23
PRIOR APPLICATION NUMBER:	60/073612
PRIOR FILING DATE:	1998-02-04
PRIOR APPLICATION NUMBER:	60/074086
PRIOR FILING DATE:	1998-02-09
PRIOR APPLICATION NUMBER:	60/074092
PRIOR FILING DATE:	1998-02-09
PRIOR APPLICATION NUMBER:	60/077791
PRIOR FILING DATE:	1998-03-12
PRIOR APPLICATION NUMBER:	60/078910
PRIOR FILING DATE:	1998-03-20
PRIOR APPLICATION NUMBER:	60/079294
PRIOR FILING DATE:	1998-03-25
PRIOR APPLICATION NUMBER:	60/079663
PRIOR FILING DATE:	1998-02-27
PRIOR APPLICATION NUMBER:	60/079728
PRIOR FILING DATE:	1998-03-27
PRIOR APPLICATION NUMBER:	60/080155
PRIOR FILING DATE:	1998-03-31
PRIOR APPLICATION NUMBER:	60/081203
PRIOR FILING DATE:	1998-04-09
PRIOR APPLICATION NUMBER:	60/081259
PRIOR FILING DATE:	1998-04-09
PRIOR APPLICATION NUMBER:	60/081555
PRIOR FILING DATE:	1998-04-14
PRIOR APPLICATION NUMBER:	60/081517
PRIOR FILING DATE:	1998-04-15
PRIOR APPLICATION NUMBER:	60/081818
PRIOR FILING DATE:	1998-04-15
PRIOR APPLICATION NUMBER:	60/082999
PRIOR FILING DATE:	1998-04-24
PRIOR APPLICATION NUMBER:	60/083222
PRIOR FILING DATE:	1998-04-28
PRIOR APPLICATION NUMBER:	60/083545
PRIOR FILING DATE:	1998-04-29
PRIOR APPLICATION NUMBER:	60/084600
PRIOR FILING DATE:	1998-05-07
PRIOR APPLICATION NUMBER:	60/084627
PRIOR FILING DATE:	1998-05-07
PRIOR APPLICATION NUMBER:	60/084637
PRIOR FILING DATE:	1998-05-07
PRIOR APPLICATION NUMBER:	60/085148
PRIOR FILING DATE:	1998-05-12
PRIOR APPLICATION NUMBER:	60/085322
PRIOR FILING DATE:	1998-05-13
PRIOR APPLICATION NUMBER:	60/085333
PRIOR FILING DATE:	1998-05-13

1	PRIOR	APPLICATION	NUMBER: 60/08583393
2	PRIOR	FILING DATE: 1998-05-13	
3	PRIOR	FILING DATE: 1998-05-13	
4	PRIOR	APPLICATION NUMBER: 60/08555797	
5	PRIOR	FILING DATE: 1998-05-15	
6	PRIOR	APPLICATION NUMBER: 60/08569797	
7	PRIOR	FILING DATE: 1998-05-15	
8	PRIOR	APPLICATION NUMBER: 60/08570404	
9	PRIOR	FILING DATE: 1998-05-15	
10	PRIOR	APPLICATION NUMBER: 60/08644144	
11	PRIOR	FILING DATE: 1998-05-22	
12	PRIOR	APPLICATION NUMBER: 60/08643030	
13	PRIOR	FILING DATE: 1998-05-22	
14	PRIOR	APPLICATION NUMBER: 60/08710606	
15	PRIOR	FILING DATE: 1998-05-28	
16	PRIOR	APPLICATION NUMBER: 60/08802626	
17	PRIOR	FILING DATE: 1998-06-04	
18	PRIOR	APPLICATION NUMBER: 60/08873030	
19	PRIOR	FILING DATE: 1998-06-10	
20	PRIOR	APPLICATION NUMBER: 60/08874141	
21	PRIOR	FILING DATE: 1998-06-10	
22	PRIOR	APPLICATION NUMBER: 60/08881010	
23	PRIOR	FILING DATE: 1998-06-10	
24	PRIOR	APPLICATION NUMBER: 60/08885656	
25	PRIOR	FILING DATE: 1998-06-11	
26	PRIOR	APPLICATION NUMBER: 60/08952323	
27	PRIOR	FILING DATE: 1998-06-17	
28	PRIOR	APPLICATION NUMBER: 60/08959999	
29	PRIOR	FILING DATE: 1998-06-17	
30	PRIOR	APPLICATION NUMBER: 60/08990707	
31	PRIOR	FILING DATE: 1998-06-18	
32	PRIOR	APPLICATION NUMBER: 60/08994747	
33	PRIOR	FILING DATE: 1998-06-19	
34	PRIOR	APPLICATION NUMBER: 60/09003494	
35	PRIOR	FILING DATE: 1998-06-23	
36	PRIOR	APPLICATION NUMBER: 60/09042292	
37	PRIOR	FILING DATE: 1998-06-24	
38	PRIOR	APPLICATION NUMBER: 60/09044545	
39	PRIOR	FILING DATE: 1998-06-24	
40	PRIOR	APPLICATION NUMBER: 60/09053898	
41	PRIOR	FILING DATE: 1998-06-24	
42	PRIOR	APPLICATION NUMBER: 60/09086363	
43	PRIOR	FILING DATE: 1998-06-26	
44	PRIOR	APPLICATION NUMBER: 60/09136636	
45	PRIOR	FILING DATE: 1998-07-01	
46	PRIOR	APPLICATION NUMBER: 60/09155191	
47	PRIOR	FILING DATE: 1998-07-02	
48	PRIOR	APPLICATION NUMBER: 60/09198282	
49	PRIOR	FILING DATE: 1998-07-07	
50	PRIOR	APPLICATION NUMBER: 60/09198282	

Query Match	8.9%;	Score 164.5;	DB 9;	Length 204;
Best Local Similarity	27.7%;	Pred. No. 4.1e-05;		
Matches 46;	Conservative 31;	Mismatches 74;	Indels 15;	Gaps 5

Qy 18 L L G V L G A V S G L S T E P P Y V N S A N K F Q A E G G V L Y P Q I G D L D L C P A R A P D G P S S - P N Y 76  
Db 8 L L G C S L A A D R A T V F W N S N P F R N E - D Y T T H V Q A N D V D I I C P Y A E - - D E S A D A A M 63  
Qy 77 E F Y L Y L V G A G Q R C E A P P A P N L L L C D S P D L - - D L A F T I K O E S P N I M G H E P R S H - 133  
Db 64 E Q I L L V L V E H E F Y D Q P Q S D Q Y R M C N C N P S A G H G E K S S E K F Q R T P T P L G N F R E G H 122  
Qy 134 D Y I I A T S D G T R E G L E S T I Q G V C I T R G M K Y L L R V G Q S P R G A N P R K 179  
Db 124 S Y Y I S K P I H O E D R - - - - C I R L V T A S G K I T T S P Q A H D N P Q E 162

RESULT 30  
US-10-121-049-288  
; Sequence 288, Application US/10121049  
; Publication No. US20030022239A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen









```

; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; FILE REFERENCE: DEX-0269
; CURRENT APPLICATION NUMBER: US/10/001,887
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,998
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: 60/252,563
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 108
; LENGTH: 638
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-887-108

```

```

Query Match 7.5%; Score 138; DB 9; Length 638;
Best Local Similarity 27.9%; Pred. No. 0.011;
Matches 68; Conservative 13; Mismatches 79; Indels 84; Gaps 12;

```

```

QY 159 RGMKTLRVGQ-----SPRGAVPRKPYSEMPMERDGAASLEPCKENTLPGDPTSNATS 213
DB 137 RLKGDNGVGGQGLPGAPGQGGAPFP--GLP-----GPAGLGKPLDGLPGAPGDKGES 189
QY 214 -----RGABGPLPPSPMPAVA--GAAG-----G 234
DB 190 GPPGVPRGREGAVGPPGPGVDGVGAAGLPGQGPBGAKGPEPTRGPPGLIGPTG 249
QY 235 LALLL-----GYAGGAGAMCWRRRRAKPSRHPG--PSFGGSLGCGGGCKGPR 286
DB 250 YGMPGLPGKGDGRGAGVGLL---GDRGEFGEDDPEBQGPQGLGPPGGLPGSAGLPGR 306
QY 287 EAPELGLALRG-----GAADPPCFHYKXVSGDYGHPIYVQDGP--P 330
DB 307 RGPPEPKGAGCGPPGPGIRGDDGPGSLAKKPGVPRERGLPGAHGP-----GPTGP 360
QY 331 QSPF 334
DB 361 KGEF 364

```

```

RESULT 40
US-09-823-240-10
; Sequence 10, Application US/09823240
; Patent No. US2002004813A1
; GENERAL INFORMATION:
; APPLICANT: James E. Gertler
; APPLICANT: Frank B. Bear
; APPLICANT: Jurgen Wehland
; APPLICANT: Joseph Loureiro
; TITLE OF INVENTION: Methods and Products for Regulating Cell
; FILE REFERENCE: M0656/7064 (HCL)
; CURRENT APPLICATION NUMBER: US/09/823,240
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/194,564
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-823-240-10

```

```

Query Match 7.1%; Score 130.5; DB 10; Length 380;
Best Local Similarity 22.8%; Pred. No. 0.022;
Matches 67; Conservative 34; Mismatches 102; Indels 91; Gaps 15;

```

```

QY 35 WMSAKRPAEGGYLYPIGIRLDDLCRRAPPPHSPHYEYKLYLVGAGRCCEA 94
DB 16 YDDGNKRWLPAG--TGQAFSRVQI-----YHNPANSFRVY-----GRKKQ- 55

```

```

QY 95 PPAENLLTCDRPDLDRTIKQSESPY-----LMGHEFRSHHDYIIATSDGR 145
DB 56 -PQQVAVINC-----ATVGVKYNQATPETHQWRDAPQVWGNFSGKEPAQFPA--GMA 107
QY 146 EGLESLOGVCULTRGMKVLIRVQSPRG-----AVPRKPYSEMPMERD--GAASLE- 197
DB 108 SALEALEGG-----GPPPPALPTWSVNPSPSEVEEQKQKQGPSEHIER 154
QY 198 -----PKENLPGDPTSN--ATSRGA--BGLPPSPMPAVA 229
DB 155 RVSNAGPPAPGAPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 214
QY 230 GAAGG-ALLLGVAGAGAMCWRRRRAKPSRHPGSGFSGSLGAGGGGM 283
DB 215 GREGGGA-----GAPGLAAIAGAKLR-KYSKQSEASGPTAKASGRSGGGGL 263

```

```

RESULT 41
US-09-925-299-979
; Sequence 979, Application US/0925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 979
; LENGTH: 674
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-299-979

```

```

Query Match 7.1%; Score 130.5; DB 10; Length 674;
Best Local Similarity 27.4%; Pred. No. 0.042;
Matches 72; Conservative 20; Mismatches 92; Indels 79; Gaps 15;

```

```

QY 141 SDGTREGLSLOGVCULTRGMKVLIRVQSP--RG--GAVPRKPYSEMPME--RDRGAH 194
DB 186 SRGR-GLPGVAGAV---GEPPLGIAGPPGARGPPAVGSPVVGAGAGAGRDGNPGN 240
QY 195 SLEPGKENLPG-----DPTSNATSRGAEGLPP-----PMPAVA 232
DB 241 DGPGRDGGGKHGKRGYGNITGVGAGAPGPPGAVGAGHGNRGRTGPGP--VGRPA 298
QY 233 GGIALLLGVAGAGAMCWRRRRAKPSRHPG--PSFGGSLGL-----GGGG 281
DB 299 GA-----VGRPGSPGQGRGDKGEGEKPRGLGLKGNHNLQGLPGIAGHNGDQGA 353
QY 282 GMP-----REARGLGIA-LRG-----GGADPPCFHYKXVSG 316
DB 354 SVGPAPRPRGPGSGPAGDGTGHPGTGVPAGIRGPGQHGPGAPGPPGPPGPGVSG 413
QY 317 ---DYGHPIYVQDPPQSPNNI 336
DB 414 GYDFGYDDGFYRADQPRASPL 436

```

```

RESULT 42
US-09-854-133-187
; Sequence 187, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Iodas, Michael J.
; APPLICANT: Mohamed, Raedoh
; APPLICANT: Henderson, Robert A.

```

```

; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 187
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-854-133-187

```

Query Match 7.0%; Score 129; DB 9; Length 595;

Best Local Similarity 27.4%; Pred. No. 0.0477; Mismatches 89; Indels 76; Gaps 12;

```

Db 163 GGAESPFGWEAGP-----REMGSPSGHGDGPRRRPRKRRGRKG 201
QY 128 EFRSHADYIIATSDGTREGLESIGGVCTTRGMKVLRYVQSPRGAVPRKPYSE---- 183
Db 202 RMGRHEAATAATATATGTGTAEAGASAPES-----QAGGPRGRA--RGPRQGRRR 254
QY 184 MPMERDGAHSLPEPKENLPDPTG---NATSRGAGPLPPEPMVAAGAGLALLL 240
Db 255 HGQRRRGPPQAREEG---PRDATTILGLTSPGEGRADQSGLPALAGAAAAHAHAIP 310
QY 241 G-----VAGAGAMCWRRRRAKPSRRHPGSGFGRGSLGLGGGGMGPREEAPGEL 293
Db 311 GAGPAAAPVGRGRGRGGRGRGRGGS---AGAGGGRGRG--GRGRGGRG----- 357
QY 294 GIALRGGAADP 305
Db 358 GGAGRGGAAGP 369

```

```

QY 294 GIALRGGAADP 305
Db 358 GGAGRGGAAGP 369

```

```

RESULT 43
US-09-738-973-187
; Sequence 187 Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodch
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Eliot, Mark
; APPLICANT: Mamion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 187
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-738-973-187

```

Query Match 7.0%; Score 129; DB 10; Length 595;

Best Local Similarity 27.4%; Pred. No. 0.0477; Mismatches 89; Indels 76; Gaps 12;

```

QY 85 GGAQG-RCEAPAPNLLTCDRPDLRFTIKQEYSPNLMGH----- 127
Db 163 GGAESPFGWEAGP-----REMGSPSGHGDGPRRRPRKRRGRKG 201
QY 128 EFRSHADYIIATSDGTREGLESIGGVCTTRGMKVLRYVQSPRGAVPRKPYSE---- 183
Db 202 RMGRHEAATAATATATGTGTAEAGASAPES-----QAGGPRGRA--RGPRQGRRR 254
QY 184 MPMERDGAHSLPEPKENLPDPTG---NATSRGAGPLPPEPMVAAGAGLALLL 240
Db 255 HGQRRRGPPQAREEG---PRDATTILGLTSPGEGRADQSGLPALAGAAAAHAHAIP 310
QY 241 G-----VAGAGAMCWRRRRAKPSRRHPGSGFGRGSLGLGGGGMGPREEAPGEL 293
Db 311 GAGPAAAPVGRGRGRGGRGRGRGGS---AGAGGGRGRG--GRGRGGRG----- 357
QY 294 GIALRGGAADP 305
Db 358 GGAGRGGAAGP 369

```

RESULT 44

```

US-09-919-497-56
; Sequence 56 Application US/09919497
; Patent No. US2002010662A1
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 1806
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (758)..(758)
; OTHER INFORMATION: Xaa = any amino acid
; NAME/KEY: UNSURE
; LOCATION: (809)..(809)
; OTHER INFORMATION: Xaa = any amino acid
US-09-919-497-56

```

Query Match 7.0%; Score 129; DB 10; Length 1806;

Best Local Similarity 29.1%; Pred. No. 0.16; Indels 80; Gaps 13;

```

QY 169 QSPRGAVPRKPYSEPMER-----DRGAHSLPEKGE-----NLRPD----- 206
Db 561 QGPRGVGPPGPTGX-PQKGRPGADGGRGWPGRPAKXDRGFDGLGRLGDKHGRERG 619
QY 207 -----PNSNATSGAGPLPPEPMVAAGAGLALLLGVAGAGAMCWRRRRAKPYSE 260
Db 620 PQGPPGPDGDMGEGDEIGPRGLPGEAGPRG-----LIGPRTPA-----PQ 665
QY 261 -----SRHPG-PGSFGRGSLGL-GGGGGMGPREEAPGELGIALRG 300
Db 666 PGMAVDPGPPKNGMPQGPBPGRGQGNPGQGLGPGPGPIGP-----PGEKPGPKPG 721
QY 301 -----GAADPPFCPHYKVGSDYGHPIYIVODGP--PQSP 333
Db 722 LAGLPADGPPGHGKGGGSEK-----ALGPPGPG 755

```

RESULT 45

```

US-09-925-302-689
; Sequence 689 Application US/09925302

```

Patent No. US20020044941A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA104  
CURRENT APPLICATION NUMBER: US/09/925,302  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05918  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 896  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 688  
LENGTH: 403  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (183)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-302-689

Query Match 6.8%; Score 126.5; DB 10; Length 403;  
Best Local Similarity 30.9%; Pred. No. 0.046;  
Matches 46; Conservative 9; Mismatches 53; Indels 41; Gaps 8;

QY 198 PCKENLPDDPSNMTS-----RGARGLPSPSPMAVAGAGLMLLLGVAGAGGAMCWR 252  
DB 168 PGAPGPPPEPEKAGXGAMGDBGATGPGPGPGVKGBA-----LQPGGARG----- 217

QY 253 RRAPKSSRRHPGPG-----GSPFGGSLGLGGGGGMPREAPPEGLI-----ALRGCG 301  
DB 218 ----KQAGTGTGPGGEGSGSKDGLIGPKETG---TKGSKDGLGLPGSKGDRGMKGDA 270

QY 302 AADPFCHYEKVSQDYHPVYIVQDGP 330  
DB 271 GVMGP--PGAQSGKDFGRP-----GPP 291

RESULT 46  
US-09-978-295A-614  
Sequence 614, Application US/09978295A  
GENERAL INFORMATION:  
APPLICANT: Ashkenazi, Avi  
APPLICANT: Baker Kevin P.  
APPLICANT: Botstein, David  
APPLICANT: Desnovers, Luc  
APPLICANT: Eaton, Dan  
APPLICANT: Ferrara, Napoleon  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gerltsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Kljavin, Ivar J.  
APPLICANT: Kuo, Sophia S.  
APPLICANT: Napier, Mary A.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Shelton, David L.  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

TITLE OF INVENTION: Acids Encoding the Same  
FILE REFERENCE: P26302P1C11  
CURRENT APPLICATION NUMBER: US/09/978,295A  
CURRENT FILING DATE: 2001-10-15  
PRIOR APPLICATION NUMBER: 09/918585  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/064249  
PRIOR FILING DATE: 1997-11-03  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066364  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/077450  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: 60/077632  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077641  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077649  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077791  
PRIOR FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: 60/078004  
PRIOR FILING DATE: 1998-03-13  
PRIOR APPLICATION NUMBER: 60/078886  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078936  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078939  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079294  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: 60/079656  
PRIOR FILING DATE: 1998-03-26  
PRIOR APPLICATION NUMBER: 60/079664  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079689  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079663  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079728  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079786  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079920  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/079923  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/080105  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080107  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080165  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080194  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080327  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080328  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080333  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080334  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/081070  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081049  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081071

;; PRIOR FILING DATE: 1998-04-08  
;; PRIOR APPLICATION NUMBER: 60/081195  
;; PRIOR FILING DATE: 1998-04-08  
;; PRIOR APPLICATION NUMBER: 60/081203  
;; PRIOR FILING DATE: 1998-04-09  
;; PRIOR APPLICATION NUMBER: 60/081229  
;; PRIOR FILING DATE: 1998-04-09  
;; PRIOR APPLICATION NUMBER: 60/081955  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/081817  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/081819  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/081952  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/081838  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/082568  
;; PRIOR FILING DATE: 1998-04-21  
;; PRIOR APPLICATION NUMBER: 60/082569  
;; PRIOR FILING DATE: 1998-04-21  
;; PRIOR APPLICATION NUMBER: 60/082704  
;; PRIOR FILING DATE: 1998-04-22  
;; PRIOR APPLICATION NUMBER: 60/082804  
;; PRIOR FILING DATE: 1998-04-22  
;; PRIOR APPLICATION NUMBER: 60/082700  
;; PRIOR FILING DATE: 1998-04-22  
;; PRIOR APPLICATION NUMBER: 60/082797  
;; PRIOR FILING DATE: 1998-04-22  
;; PRIOR APPLICATION NUMBER: 60/082796  
;; PRIOR FILING DATE: 1998-04-23  
;; PRIOR APPLICATION NUMBER: 60/083336  
;; PRIOR FILING DATE: 1998-04-27  
;; PRIOR APPLICATION NUMBER: 60/083322  
;; PRIOR FILING DATE: 1998-04-28  
;; PRIOR APPLICATION NUMBER: 60/083392  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083495  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083496  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083499  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083545  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083554  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083558  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083559  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083500  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083742  
;; PRIOR FILING DATE: 1998-04-30  
;; PRIOR APPLICATION NUMBER: 60/084366  
;; PRIOR FILING DATE: 1998-05-05  
;; PRIOR APPLICATION NUMBER: 60/084414  
;; PRIOR FILING DATE: 1998-05-06  
;; PRIOR APPLICATION NUMBER: 60/084441  
;; PRIOR FILING DATE: 1998-05-06  
;; PRIOR APPLICATION NUMBER: 60/084637  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084639  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084640  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084598  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084600  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084627  
;; PRIOR FILING DATE: 1998-05-07

;; PRIOR APPLICATION NUMBER: 60/084643  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/085339  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085338  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085323  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085582  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085700  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085689  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085579  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085580  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085573  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085704  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085697

Query Match 6.8%; Score 126.5; DB 9; Length 520;  
Best Local Similarity 30.3%; Pred. No. 0.061;  
Matches 46; Conservative 9; Mismatches 53; Indels 41; Gaps 8;

Qy 198 PKNENLPGDPTSNAT-----RGAEGPLPPSMYAVGAGGJALLLGVAGAGGACMR 252  
Db 165 FGAPGPPPPAEKAKAGMGDGTGSPGPPGVGBG-----LQGQAGP----- 214  
Qy 253 RRRAPBSRRHPG---GSFGRGSLGCGGGMGRAPBGLG-----ALRGCG 301  
Db 215 ---KQANTGPPGQKKGSKGQGGGLGPKGKG---TKKEXDGLGPKSKGRKMGDA 267  
Qy 302 AADPPFCHYEKSGDYGHVYIVQDGP 330  
Db 268 GWMGP--FGAQSGKDPGR-----GPF 288

RESULT 47  
US-09-978-697-614  
; Sequence 614, Application US/09978697  
; Patent No. US20020169284A1  
; GENERAL INFORMATION:  
; APPLICANT: Ashkenazi, Avi  
; APPLICANT: Baker Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Denoyez, Luc  
; APPLICANT: Baton, Dan  
; APPLICANT: Ferrara, Napoleon  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerlitsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, J. Christopher  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth J  
; APPLICANT: Kijavlin, Ivar J.  
; APPLICANT: Kuo, Sophia S.  
; APPLICANT: Napier, Mary A.  
; APPLICANT: Pan, James;  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Shelton, David L.  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Williams, P. Mckey  
; APPLICANT: Wood, William I.

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
ACIDS  
FILE REFERENCE: P2630PIC27  
CURRENT FILING DATE: 2001-10-16  
PRIOR APPLICATION NUMBER: 09/918585  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/064249  
PRIOR FILING DATE: 1997-11-03  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066364  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/077450  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: 60/077632  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077641  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077649  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077791  
PRIOR FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: 60/078004  
PRIOR FILING DATE: 1998-03-13  
PRIOR APPLICATION NUMBER: 60/078866  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078936  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078939  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079294  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: 60/079656  
PRIOR FILING DATE: 1998-03-26  
PRIOR APPLICATION NUMBER: 60/079664  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079689  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079663  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079728  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079786  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079920  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/079923  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/080105  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080107  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080165  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080194  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080327  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080328  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080333  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080334  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/081070  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081049  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081071  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081195  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081203  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081229  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081955  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081817  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081819  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081952  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081838  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082568  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082569  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082704  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082804  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082700  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082797  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082796  
PRIOR FILING DATE: 1998-04-23  
PRIOR APPLICATION NUMBER: 60/083336  
PRIOR FILING DATE: 1998-04-27  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/083392  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083495  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083496  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083499  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083545  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083554  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083558  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083559  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083500  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083742  
PRIOR FILING DATE: 1998-04-30  
PRIOR APPLICATION NUMBER: 60/084366  
PRIOR FILING DATE: 1998-05-05  
PRIOR APPLICATION NUMBER: 60/084414  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 60/084441  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 60/084637  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084639  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084640  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084598  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084600  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084627

;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084643  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/085339  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085338  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085323  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085582  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085700  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085689  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085579  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085580  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085573  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085704  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085697

Query Match 6.8%; Score 126.5; DB 9; Length 520;  
Best Local Similarity 30.9%; Pred. No. 0.061;  
Matches 46; Conservative 9; Mismatches 53; Indels 41; Gaps 8;

QY 198 PGKENTLPGDPTSNAT-----RGAEGPLPPSPMPAYAGAGLALLLVAGAGGAMCNR 252  
DB 165 PGAPGPPGPAEKAKGAKGAMRDGATGSPGQGPVYGEAG-----LQGPQAGF----- 214  
QY 253 RRRAKPSESHPGP---GSGFRGSLGLGGGGMGPRBAEPGLGI-----ALRGGG 301  
DB 215 ---KQANTGTPEHQSGKSGKDGGLIGPKGTG---TKGKBDLGLPSGKDRMKGDA 267  
QY 302 AADPFCHYKXVSGDYGHFVTVQDGP 330  
DB 268 GVMGP--PGAQSGKDGFRP-----GPP 288

RESULT 48  
US-09-978-192A-614  
Sequence 614, Application US/09978192A  
Patent No. US20020177553A1  
GENERAL INFORMATION:  
APPLICANT: Ashkenazi, Avi  
APPLICANT: Baker, Kevin P.  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan  
APPLICANT: Ferrara, Napoleon  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gottlieb, Mary E.  
APPLICANT: Grimsdahl, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Guirney, Austin L.  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Kijavitt, Ivor J.  
APPLICANT: Kuo, Sophia S.  
APPLICANT: Napier, Mary A.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Shelton, David L.  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tuma, Daniel  
APPLICANT: Williams, P. Mickey

;; APPLICANT: Wood, William I.  
;; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
;; FILE REFERENCE: P2630PIC9  
;; CURRENT APPLICATION NUMBER: US/09/978,192A  
;; CURRENT FILING DATE: 2001-10-15  
;; PRIOR APPLICATION NUMBER: 09/918595  
;; PRIOR FILING DATE: 2001-07-30  
;; PRIOR APPLICATION NUMBER: 60/062250  
;; PRIOR FILING DATE: 1997-10-17  
;; PRIOR APPLICATION NUMBER: 60/064249  
;; PRIOR FILING DATE: 1997-11-03  
;; PRIOR APPLICATION NUMBER: 60/065311  
;; PRIOR FILING DATE: 1997-11-13  
;; PRIOR APPLICATION NUMBER: 60/066364  
;; PRIOR FILING DATE: 1997-11-21  
;; PRIOR APPLICATION NUMBER: 60/077450  
;; PRIOR FILING DATE: 1998-03-10  
;; PRIOR APPLICATION NUMBER: 60/077632  
;; PRIOR FILING DATE: 1998-03-11  
;; PRIOR APPLICATION NUMBER: 60/077641  
;; PRIOR FILING DATE: 1998-03-11  
;; PRIOR APPLICATION NUMBER: 60/077649  
;; PRIOR FILING DATE: 1998-03-11  
;; PRIOR APPLICATION NUMBER: 60/077791  
;; PRIOR FILING DATE: 1998-03-12  
;; PRIOR APPLICATION NUMBER: 60/078004  
;; PRIOR FILING DATE: 1998-03-13  
;; PRIOR APPLICATION NUMBER: 60/078086  
;; PRIOR FILING DATE: 1998-03-20  
;; PRIOR APPLICATION NUMBER: 60/078936  
;; PRIOR FILING DATE: 1998-03-20  
;; PRIOR APPLICATION NUMBER: 60/078910  
;; PRIOR FILING DATE: 1998-03-20  
;; PRIOR APPLICATION NUMBER: 60/078939  
;; PRIOR FILING DATE: 1998-03-20  
;; PRIOR APPLICATION NUMBER: 60/079294  
;; PRIOR FILING DATE: 1998-03-25  
;; PRIOR APPLICATION NUMBER: 60/079656  
;; PRIOR FILING DATE: 1998-03-26  
;; PRIOR APPLICATION NUMBER: 60/079664  
;; PRIOR FILING DATE: 1998-03-27  
;; PRIOR APPLICATION NUMBER: 60/079689  
;; PRIOR FILING DATE: 1998-03-27  
;; PRIOR APPLICATION NUMBER: 60/079663  
;; PRIOR FILING DATE: 1998-03-27  
;; PRIOR APPLICATION NUMBER: 60/079728  
;; PRIOR FILING DATE: 1998-03-27  
;; PRIOR APPLICATION NUMBER: 60/079786  
;; PRIOR FILING DATE: 1998-03-27  
;; PRIOR APPLICATION NUMBER: 60/079920  
;; PRIOR FILING DATE: 1998-03-30  
;; PRIOR APPLICATION NUMBER: 60/079923  
;; PRIOR FILING DATE: 1998-03-30  
;; PRIOR APPLICATION NUMBER: 60/080105  
;; PRIOR FILING DATE: 1998-03-31  
;; PRIOR APPLICATION NUMBER: 60/080107  
;; PRIOR FILING DATE: 1998-03-31  
;; PRIOR APPLICATION NUMBER: 60/080165  
;; PRIOR FILING DATE: 1998-03-31  
;; PRIOR APPLICATION NUMBER: 60/080194  
;; PRIOR FILING DATE: 1998-03-31  
;; PRIOR APPLICATION NUMBER: 60/080327  
;; PRIOR FILING DATE: 1998-04-01  
;; PRIOR APPLICATION NUMBER: 60/080328  
;; PRIOR FILING DATE: 1998-04-01  
;; PRIOR APPLICATION NUMBER: 60/080333  
;; PRIOR FILING DATE: 1998-04-01  
;; PRIOR APPLICATION NUMBER: 60/080334  
;; PRIOR FILING DATE: 1998-04-01  
;; PRIOR APPLICATION NUMBER: 60/081070  
;; PRIOR FILING DATE: 1998-04-08  
;; PRIOR APPLICATION NUMBER: 60/081049



PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081071  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081195  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081203  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081229  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081955  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081817  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081819  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081952  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081838  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082568  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082569  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082704  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082804  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082700  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082797  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082796  
PRIOR FILING DATE: 1998-04-23  
PRIOR APPLICATION NUMBER: 60/083336  
PRIOR FILING DATE: 1998-04-27  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/083392  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083495  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083496  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083499  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083545  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083554  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083558  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083559  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083500  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083742  
PRIOR FILING DATE: 1998-04-30  
PRIOR APPLICATION NUMBER: 60/084366  
PRIOR FILING DATE: 1998-05-05  
PRIOR APPLICATION NUMBER: 60/084414  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 60/084441  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 60/084637  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084639  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084640  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084598  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084600  
PRIOR FILING DATE: 1998-05-07

PRIOR APPLICATION NUMBER: 60/084627  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084643  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/085339  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085338  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085323  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085582  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085700  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085689  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085579  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085580  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085573  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085704  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085697

Query Match 6.8%; Score 126.5; DB 9; Length 520;  
Best Local Similarity 30.9%; Pred. No. 0.061;  
Matches 46; Conservative 9; Mismatches 53; Indels 41; Gaps 8;

QY 198 RGENLPDPTSNATSGAEGPLPMPVAVAGAGLALLGVAGAGAMCR 252  
DB 165 PGAPPPPPAEKAGKANGRDATPSGPDPVKGAG-----LQGPQAPG----- 214  
QY 253 RRAKPSERHPGP-----GSFGRGSLGLGGGKGPRRAEPGELGI-----ALRGG 301  
DB 215 ----KQAGTGPQGPQGEKSKGPGGGLGPKGTG---TKGERGDLGLPSKKGDRGAKGDA 267  
QY 302 AADPFCHPEKSGDYGHPVYVDDGP 330  
DB 268 GVMGP--PGAQSKGDFGRP-----GFP 288

RESULT 49  
US-09-999-832A-614  
Sequence 614, Application US/0999832A  
Publication No. US20020192706A1  
GENERAL INFORMATION:  
APPLICANT: Ashkenazi, Avi  
APPLICANT: Baker Kevin P.  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan  
APPLICANT: Ferrara, Napoleon  
APPLICANT: Filvaroli, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gertsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Guiney, Austin L.  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Kijavlin, Ivar J.  
APPLICANT: Kuo, Sophia S.  
APPLICANT: Napier, Mary A.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Shelton, David L.  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel

APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: P2630PLC63  
CURRENT APPLICATION NUMBER: US/09/999,832A  
CURRENT FILING DATE: 2001-10-24  
PRIOR APPLICATION NUMBER: 09/918585  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/064249  
PRIOR FILING DATE: 1997-11-03  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066364  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/077450  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: 60/077632  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077641  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077649  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077791  
PRIOR FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: 60/078004  
PRIOR FILING DATE: 1998-03-13  
PRIOR APPLICATION NUMBER: 60/078886  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078936  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078939  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079294  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: 60/079656  
PRIOR FILING DATE: 1998-03-26  
PRIOR APPLICATION NUMBER: 60/079664  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079689  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079663  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079728  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079786  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079920  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/079923  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/080105  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080107  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080165  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080194  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080327  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080328  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080333  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080334  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/081070  
PRIOR FILING DATE: 1998-04-08

PRIOR APPLICATION NUMBER: 60/081049  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081071  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081195  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081203  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081229  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081955  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081817  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081819  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081952  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081838  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082568  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082569  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082704  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082804  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082700  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082797  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082796  
PRIOR FILING DATE: 1998-04-23  
PRIOR APPLICATION NUMBER: 60/083336  
PRIOR FILING DATE: 1998-04-27  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/083392  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083495  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083496  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083499  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083545  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083554  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083558  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083559  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083500  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083742  
PRIOR FILING DATE: 1998-04-30  
PRIOR APPLICATION NUMBER: 60/084366  
PRIOR FILING DATE: 1998-05-05  
PRIOR APPLICATION NUMBER: 60/084414  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 60/084441  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 60/084637  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084639  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084640  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084598  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084600

PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084627  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084643  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/085339  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085338  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085323  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085582  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085700  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085689  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085579  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085580  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085573  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085704  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085697

Query Match 6.8%; Score 126.5; DB 9; Length 520;  
Best Local Similarity 30.9%; Pred. No. 0.061;  
Matches 46; Conservative 9; Mismatches 53; Indels 41; Gaps 8;

QY 198 PCKENIPGPTSNATS-----RGARGLPPEPMVAAGALILIGVAGAGAMCMR 252  
DB 165 PAAPGPGPAKAKAGKAGMRDAGTSGPGPGPVKKBAG-----LGGPGAPG----- 214  
QY 253 RRRAPSESRHPP-----GSFGRGSLIGGGGGMGPRAEPGELGI-----ALRGGG 301  
DB 215 ----KQAGTGTGPGGEEKSKGDDGLIGFKBTG---TKGEKGLDGLGSKGDKMGKDA 267  
QY 302 AADPPCPHYEVSGDYGHPTVVDGPP 330  
DB 268 GVMGP--PGAQSKGDFGRP-----GPP 288

RESULT 50  
US-03-978-189-614  
Sequence 614, Application US/09978189  
Publication No. US20030004102A1  
GENERAL INFORMATION:  
APPLICANT: Ashkenazi, Avi  
APPLICANT: Baker Kevin P.  
APPLICANT: Botstein, David  
APPLICANT: Deemeyers, Luc  
APPLICANT: Baton, Dan  
APPLICANT: Ferrara, Napoleon  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gertsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Kljavin, Ivar J.  
APPLICANT: Kuo, Sophia S.  
APPLICANT: Napier, Mary A.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Shelton, David L.  
APPLICANT: Stewart, Timothy A.

APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: Secured and Transmembrane Polypeptides and Nucleic  
Acide Encoding the Same  
FILE REFERENCE: P2630P1C7  
CURRENT APPLICATION NUMBER: US/09/978,189  
CURRENT FILING DATE: 2001-10-15  
PRIOR APPLICATION NUMBER: 09/918585  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/064249  
PRIOR FILING DATE: 1997-11-03  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066364  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/077450  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: 60/07632  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/07641  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/07649  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/07791  
PRIOR FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: 60/078004  
PRIOR FILING DATE: 1998-03-13  
PRIOR APPLICATION NUMBER: 60/078886  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078936  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078939  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079294  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: 60/079656  
PRIOR FILING DATE: 1998-03-26  
PRIOR APPLICATION NUMBER: 60/079664  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079689  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079663  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079728  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079786  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079920  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/079923  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/080105  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080107  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080165  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080194  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080327  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080328  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080333  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080334  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/081070

;; PRIOR FILING DATE: 1998-04-08  
;; PRIOR APPLICATION NUMBER: 60/081049  
;; PRIOR FILING DATE: 1998-04-08  
;; PRIOR APPLICATION NUMBER: 60/081071  
;; PRIOR FILING DATE: 1998-04-08  
;; PRIOR APPLICATION NUMBER: 60/081195  
;; PRIOR FILING DATE: 1998-04-08  
;; PRIOR APPLICATION NUMBER: 60/081203  
;; PRIOR FILING DATE: 1998-04-09  
;; PRIOR APPLICATION NUMBER: 60/081229  
;; PRIOR FILING DATE: 1998-04-09  
;; PRIOR APPLICATION NUMBER: 60/081955  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/081817  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/081819  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/081952  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/081838  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/082568  
;; PRIOR FILING DATE: 1998-04-21  
;; PRIOR APPLICATION NUMBER: 60/082569  
;; PRIOR FILING DATE: 1998-04-21  
;; PRIOR APPLICATION NUMBER: 60/082704  
;; PRIOR FILING DATE: 1998-04-22  
;; PRIOR APPLICATION NUMBER: 60/082804  
;; PRIOR FILING DATE: 1998-04-22  
;; PRIOR APPLICATION NUMBER: 60/082700  
;; PRIOR FILING DATE: 1998-04-22  
;; PRIOR APPLICATION NUMBER: 60/082797  
;; PRIOR FILING DATE: 1998-04-22  
;; PRIOR APPLICATION NUMBER: 60/082796  
;; PRIOR FILING DATE: 1998-04-23  
;; PRIOR APPLICATION NUMBER: 60/083336  
;; PRIOR FILING DATE: 1998-04-27  
;; PRIOR APPLICATION NUMBER: 60/083322  
;; PRIOR FILING DATE: 1998-04-28  
;; PRIOR APPLICATION NUMBER: 60/083392  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083495  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083496  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083499  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083545  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083554  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083558  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083559  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083500  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083742  
;; PRIOR FILING DATE: 1998-04-30  
;; PRIOR APPLICATION NUMBER: 60/084366  
;; PRIOR FILING DATE: 1998-05-05  
;; PRIOR APPLICATION NUMBER: 60/084414  
;; PRIOR FILING DATE: 1998-05-06  
;; PRIOR APPLICATION NUMBER: 60/084441  
;; PRIOR FILING DATE: 1998-05-06  
;; PRIOR APPLICATION NUMBER: 60/084637  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084639  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084640  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084598  
;; PRIOR FILING DATE: 1998-05-07

;; PRIOR APPLICATION NUMBER: 60/084600  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084627  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084643  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/085339  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085338  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085323  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085582  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085700  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085689  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085579  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085580  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085573  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085704  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085697

Query Match 6.8%; Score 126.5; DB 9; Length 520;  
Best Local Similarity 30.9%; Pred. No. 0.061;  
Matches 46; Conservative 9; Mismatches 53; Indels 41; Gaps 8;

QY 198 PKXENLPGDPTSNATG-----RGAEGPLPPSPMRAVGAAGLALLLGVAGAGGAMCWR 252  
DB 165 PGADGPPGPPKAKAKAMGMDGATGSPGQGPVGXGNG-----LQGFQAGP----- 214  
QY 253 RRAKPSRSRHPGP-----GSFGRGSLGLGCGGGMGPREAPEGLGI-----ALRGCG 301  
DB 215 ----KQATGTGPGQKKGSGKGGGLGIPKGTG---TKEXKDLGLPGSKGDRGMWGDA 267  
QY 302 AADPPCPHIEKVSQDYGHPIYIVQDGP 330  
DB 268 GVMGP--FGAGSKXDRGP-----GPP 288

Search completed: February 11, 2003, 12:13:17  
Job time : 20.5409 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2003, 11:47:56 / Search time 25.2327 Seconds  
(without alignments)  
2776.398 Million cell updates/sec

Title: US-10-021-121-4

Perfect score: 1850  
Sequence: 1 MGFPHSGPGVRRVALLILG.....PYIVDPGPQSPNIVYKV 340

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 50 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP\_invertebrate:\*  
6: SP\_mammal:\*  
7: SP\_mhc:\*  
8: SP\_organelle:\*  
9: SP\_phase:\*  
10: SP\_plant:\*  
11: SP\_rodent:\*  
12: SP\_virus:\*  
13: SP\_vertebrate:\*  
14: SP\_unclassified:\*  
15: SP\_virus:\*  
16: SP\_bacteriap:\*  
17: SP\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1559	84.3	285	4 Q8TBH7	Q8TBH7 homo sapien
2	923	49.9	331	13 Q90Z31	Q90Z31 brachydanio
3	919	49.7	327	13 Q9PT69	Q9PT69 xenopus lae
4	629.5	34.0	333	13 Q9PU04	Q9PU04 gallus gall
5	616	33.3	341	13 Q90Z32	Q90Z32 brachydanio
6	605.5	32.7	334	13 Q90Z32	Q90Z32 brachydanio
7	340	18.4	205	13 Q9W6H9	Q9W6H9 xenopus lae
8	214.5	11.6	237	5 Q9U3M2	Q9U3M2 caenorhabdi
9	198.5	10.7	279	5 Q9U474	Q9U474 caenorhabdi
10	182	9.8	652	5 Q9V4E1	Q9V4E1 drosophila
11	178.5	9.6	202	13 Q98T21	Q98T21 gallus gall
12	167	9.0	205	11 Q9D7K8	Q9D7K8 mus musculu
13	163.5	8.8	206	11 Q9CZS8	Q9CZS8 mus musculu
14	156	8.4	675	6 Q9N178	Q9N178 sus scrofa
15	155	8.4	229	13 Q9J431	Q9J431 brachydanio
16	153.5	8.3	1691	11 Q9ESQ2	Q9ESQ2 mus musculu

17	151.5	8.2	1447	13 Q9TB91	Q9TB91 xenopus lae
18	149	8.1	325	5 Q17036	Q17036 caenorhabdi
19	149	8.1	569	5 Q17208	Q17208 bombyx mori
20	148.5	8.0	316	5 Q19111	Q19111 caenorhabdi
21	146.5	7.9	1140	11 Q61434	Q61434 mus musculu
22	146.5	7.9	1491	13 Q91718	Q91718 xenopus lae
23	146.5	7.9	1774	11 Q62001	Q62001 mus musculu
24	146	7.9	675	13 Q90800	Q90800 gallus gall
25	146	7.9	1669	11 Q9CZS0	Q9CZS0 mus musculu
26	145.5	7.9	305	5 Q17805	Q17805 caenorhabdi
27	145.5	7.9	308	5 Q94620	Q94620 meloiodogryne
28	145	7.8	680	11 Q9D0D2	Q9D0D2 mus musculu
29	145	7.8	1745	4 Q9NZ06	Q9NZ06 homo sapien
30	144.5	7.8	309	5 Q25466	Q25466 meloiodogryne
31	144.5	7.8	1347	4 Q960R3	Q960R3 homo sapien
32	144.5	7.8	1420	13 Q90M37	Q90M37 gallus gall
33	143.5	7.8	775	16 Q9R342	Q9R342 streptomyce
34	143	7.7	1747	5 Q26640	Q26640 strongyloce
35	142	7.7	744	4 Q96D07	Q96D07 homo sapien
36	142	7.7	1835	13 Q91AV4	Q91AV4 gallus gall
37	141	7.6	284	5 Q25582	Q25582 teladorsagi
38	141	7.6	284	5 Q25581	Q25581 teladorsagi
39	141	7.6	319	5 Q17038	Q17038 caenorhabdi
40	141	7.6	809	13 Q93485	Q93485 oncorhynch
41	141	7.6	1449	13 Q910C0	Q910C0 oncorhynch
42	140.5	7.6	219	13 Q90YC5	Q90YC5 brachydanio
43	140.5	7.6	305	5 P91274	P91274 caenorhabdi
44	140	7.6	304	5 Q9XUE9	Q9XUE9 caenorhabdi
45	140	7.6	1378	5 Q97405	Q97405 haliotis di
46	139.5	7.5	1450	13 Q9Y1B4	Q9Y1B4 cynops pyrr
47	139.5	7.5	1690	4 Q9UWG6	Q9UWG6 homo sapien
48	139.5	7.5	1691	4 Q9Y4L4	Q9Y4L4 homo sapien
49	139	7.5	705	4 Q8TEJ5	Q8TEJ5 homo sapien
50	139	7.5	744	11 Q9DZV4	Q9DZV4 mus musculu

## ALIGNMENTS

RESULT 1	
Q8TBH7	PRELIMINARY; PRT; 285 AA.
AC Q8TBH7	01-JUN-2002 (TREMBL) 21, Created
DT 01-JUN-2002 (TREMBL) 21, Last sequence update)	
DT 01-JUN-2002 (TREMBL) 21, Last annotation update)	
DE Similar to ephrin B3 (Fragment).	
OS Homo sapiens (Human)	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
OX NCBI_TaxID=9606;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC TISSUE=BRAIN;	
RA Strauberg R;	
BL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.	
DR EMBL; BC022499; AA022499.1; -	
FT NON_TER	
FT SEQUENCE	285 AA; 30159 MW; 08D8ACE7142469B2 CRC64;
Query Match	84.3%; Score 1559; DB 4; Length 285;
Best Local Similarity	100.0%; Pred. No. 3.3e-112;
Matches	285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	56 DRLDLCRRARPQPHSPNYEYKYLIVGAGQRCRCAPPAENLLITCDRPDIDLRFTI 115
DB	1 DRLDLCRRARPQPHSPNYEYKYLIVGAGQRCRCAPPAENLLITCDRPDIDLRFTI 60
QY	116 KFOEESPILMGHEFRSHDYIITISDGTREGLESILGGVLTREGMKVTLRVGQSPRGA 175
DB	61 KFOEESPILMGHEFRSHDYIITISDGTREGLESILGGVLTREGMKVTLRVGQSPRGA 120
QY	176 VPRKVSFEMPERDRGAHSLPEKKNLPQDPTSNATSRGAEGLPPSPMPAVAGAGGL 235

DB 121 VERKVSSEMPERDRGAASLEPGKENTLPGPTSNATRGEGPLPPSPMAVGAAGL 180  
 QY 236 ALLLGVAGAGAMCWRRRRAKPSRRHPGPGSFGRGSLGLGGGGGMPREAPGELGI 295  
 DB 181 ALLLGVAGAGAMCWRRRRAKPSRRHPGPGSFGRGSLGLGGGGGMPREAPGELGI 240  
 QY 296 ALRGGAADPPPCPHYEKVSGDYGHPIYIVODGPPQSPNNIYKV 340  
 DB 241 ALRGGAADPPPCPHYEKVSGDYGHPIYIVODGPPQSPNNIYKV 285

## RESULT 2

Q90231 PRELIMINARY; PRT; 331 AA.  
 ID 090231;  
 AC 090231;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Ephrin B3  
 OS Brachydanio rerio (Zebrafish) (Zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OC NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21290827; PubMed=11397014; Chen J.N., Goldstein N.B.,  
 RA Chan J., Mably J.D., Serluca F.C., Chen J.N., Goldstein N.B.,  
 RA Thomas M.C., Cleary J.A., Brennan C., Fishman M.C., Roberts T.M.;  
 RT "Morphogenesis of prechordal plate and notochord requires intact  
 RT eph/ephrin b signaling.";  
 RL Dev. Biol. 234:470-482(2001).  
 DR EMBL; AF375227; AAK64277.1;  
 DR InterPro; IPR001799; Ephrin.  
 DR Pfam; PF00812; Ephrin; 1.  
 DR ProDom; PD002533; Ephrin; 1.  
 DR PROSITE; PS01299; EPHRIN; UNKNOWN\_1.  
 SQ SEQUENCE 331 AA; 35638 MW; 6A5EACD509A09818 CRC64;

Query Match 49.9%; Score 923; DB 13; Length 331;  
 Best Local Similarity 54.8%; Pred. No. 3.1e-63;  
 Matches 190; Conservative 43; Mismatches 74; Indels 40; Gaps 10;

QY 10 GVRVALLLGLVGLSLEPYVWNSANKRFQAEQVLYVPOIGRDLDCRRAAPPG 69  
 DB 9 GGLILLPLVLDLG-ITATNMEPIYNSLNKRFSDKGYVLYPQIGRDLDCSSDPG 67  
 QY 70 PHSPNVEFYKLYVGG-AQGRCEAPAPAPNLLTCRDPDLDFRTIKFQESPNLWGH 128  
 DB 68 PAPADVEYKLYLVSSREQADRCVGTAPNLLTCRDPNMDMRTTIFQESPNLWGH 127  
 QY 129 FSHHDYIATSDGTRBGLSLGGVCLTRGMKYLTVVGSPRG-GAVPRKPYSEMPME 187  
 DB 128 FTTNHDYIATSDGTRBGLSLGGVCLTRGMKYLTVVGSPRG-GAVPRKPYSEMPME 182  
 QY 188 RDRGAHSLPEKENVLPDPTSNAT-----SRGAEGLPPSPMPAVAGAAGLALL 239  
 DB 183 -----AGRINNPNPGTGNSTHPOIPPRGSGENGEPPLPASNIATVIAAGAGSAFL 232  
 QY 240 LGVAGAGAMCWRRRRAKPSRRHPGPGSFG-----RGSLGCGGGGMPREAPGEL 293  
 DB 233 L-VTAVTICVCTRRRAKPSRRHP-PLSLSLTPKRGCGGGGAGGANNNG---SESDI 287  
 QY 294 GIALRGGAADPPPCPHYEKVSGDYGHPIYIVODGPPQSPNNIYKV 340  
 DB 288 IIPLR---TSDAYCPHYEKVSGDYGHPIYIVODGPPQSPNNIYKV 331

RESULT 3  
 Q9PT69 PRELIMINARY; PRT; 327 AA.  
 ID 09PT69;  
 AC 09PT69;

DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Ephrin-B3 precursor.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Fipidae; Fipidae;  
 OC Xenopodinae; Xenopus.  
 OC NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=EMBRYO;  
 RX MEDLINE=20099673; PubMed=10633856;  
 RA Heibling P.M., Saulnier D.M.E., Robinson V., Christiansen J.H.,  
 RA Wilkinson D.G., Brandt A.W.;  
 RT "Comparative analysis of embryonic gene expression defines potential  
 RT interaction sites for Xenopus EphB4 receptors with ephrin-B ligands.";  
 RL Dev. Dyn. 216:361-373(1999).  
 DR EMBL; AJ236866; CAB5511.1;  
 DR InterPro; IPR002086; Aldehyde\_dehydr.  
 DR InterPro; IPR001799; Ephrin.  
 DR Pfam; PF00812; Ephrin; 1.  
 DR ProDom; PD002533; Ephrin; 1.  
 DR PROSITE; PS00687; ALDEHYDE\_DEHYDR\_GLU; UNKNOWN\_1.  
 DR PROSITE; PS01299; EPHRIN; 1.  
 KW Signal.  
 FT SIGNAL.  
 SQ SEQUENCE 327 AA; 35913 MW; 4BB0FA39D4C22DCD CRC64;

Query Match 49.7%; Score 919; DB 13; Length 327;  
 Best Local Similarity 60.4%; Pred. No. 6.1e-63;  
 Matches 192; Conservative 30; Mismatches 82; Indels 14; Gaps 8;

QY 25 VGLSLSEPYVWNSANKRFQAEQVLYVPOIGRDLDCRRAAPPGPHSPNTEFYKLYV 84  
 DB 22 ISALSLDPTVWNSNKRFPDTEGYVLYPQIGRDLDCRSPQGFSSPYEYKLYV 81  
 QY 85 GGAQG-RCEAPAPNLLTCRDPDLDFRTIKFQESPNLWGHFRSHHDYIATSDG 143  
 DB 82 GTKEWSSCSILRTPLLLTCRDPGDLRTTIKQEFSPNLWGHFGQSRDYIATSDG 141  
 QY 144 TRBGLSLGGVCLTRGMKYLTVVGSPRGAVPRKPYSEMPERDRGAASL-EPGKN 202  
 DB 142 TMDGIRTLGGVCTGKMTLVKVGSPGATPPRRPS---AGKDSGLSPVPNDIPN 198  
 QY 203 LPDPTSNATRGAEGLPPSPMPAVAGAAGLALLLGVAGAGAMCWRRRRAKPSER 262  
 DB 199 V-GEISGNATKGENGFLLISHVPLVAGAAGLALLL-VFAYGVCHRRQAKHSDIR 256  
 QY 263 HPGPFGRGSLGLGGGGGMPREAPGELIARGGGAADPPPCPHYEKVSGDYGHV 322  
 DB 257 HP-PLSLGSLTSPKRGNN-----NGHEPDIIMPLRPSAAG-AFCPHYEKVSGDYGHV 309  
 QY 323 YIVODGPPQSPNNIYKV 340  
 DB 310 YIVODMASQSPANNIYKV 327

RESULT 4  
 Q9PU24 PRELIMINARY; PRT; 333 AA.  
 ID 09PU24;  
 AC 09PU24;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE Ephrin-B2 precursor.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OC NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Menzel P., Pasquale E.B.;  
 RT "Coding sequence of chicken ephrin-B2";  
 RL Submitted (AUG-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF180729; AAD53948.1; -  
 DR InterPro; IPR001799; Ephrin.  
 DR Pfam; PF00812; Ephrin.1.  
 DR ProDom; PD002533; Ephrin.1.  
 DR PROSITE; PS01239; EPHRIN; 1.  
 KW Signal.  
 FT SIGNAL.  
 FT CHAIN 28 333 EPHRIN-B2.  
 SQ SEQUENCE 333 AA; 36761 MW; 4C28E8CB211B783 CRC64;

Query Match 34.0%; Score 629.5; DB 13; Length 333;  
 Best Local Similarity 40.5%; Pred. No. 1.1e-40;  
 Matches 135; Conservative 52; Mismatches 125; Indels 21; Gaps 6;

QY 14 GALLILGVLIVGSLSPYVNSANKRFQAEQGYLYLPQIGRLDLCPRAPPSPHSS 73  
 DB 16 GALLVLRRLAKSIVLDPIYVNSNPKFLPGQGLVLYPGIDKLDIICPKV---DSKTV 72  
 QY 74 PNEYFYLYLVGAQGRCAAPAPNLLTCDPDLRTTFQGYSPNLMGHERSH 133  
 DB 73 GQYEVYVWVDKDQADSCAIRKNTPELLNCARPDDQVKTTFQFSPNLMGLEFQKX 132  
 QY 134 DYLIATSDGTREGLESLQGVCLTRGMKVLIVGQSPRGAVPRKPVSEMPER----- 188  
 DB 133 DYVISTNSGSLGANNQEGVQCTKMKLIMVGDSPNAGLP---STDPIKREQDA 189  
 QY 189 -DRGAHSLPEQENLPDPTSNATRGAEGLPPSPMPAVAGAGLALLLVAGAGG 247  
 DB 190 GTNGKSTTSPFYKDSGSSSTDG--SKAGHSILGSEVALFAGIASGCTIFVIITLVV 247  
 QY 248 AMCWRRRAKPSRSRHPGSGFRGSLGCGGGMGPREAREGELIALRGGAADPF 307  
 DB 248 LLKLRKRKRKSPQHTTLLSTLATPKRSGNN---NGSEPSDIIIFLR---TADSVF 300  
 QY 308 CPYKESGDIYGVYIVDPPQSPSPNITYKV 340  
 DB 301 CPYKESGDIYGVYIVDPPQSPSPNITYKV 333

RESULT 5  
 ID Q90233 PRELIMINARY; PRT; 341 AA.  
 AC Q90233;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
 DE Ephrin B1.  
 OS Brachydanio rerio (Zebrafish) (Zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OC NCBI\_TaxID=7955;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=21290827; PubMed=11397014;  
 RA Chan J., Mably J.D., Serluca F.C., Chen J.N., Goldstein N.B.,  
 RA Thomas M.C., Cleary J.A., Brennan C., Fishman M.C., Roberts T.M.;  
 RT "Morphogenesis of prechordal plate and notochord requires intact  
 RT eph/ephrin b signaling."  
 RL Dev. Biol. 234:470-482(2001).  
 RL EMBL; AF375224; AAK64274.1; -  
 DR InterPro; IPR001799; Ephrin.  
 DR Pfam; PF00812; Ephrin.1.  
 DR ProDom; PD002533; Ephrin.1.  
 DR PROSITE; PS01239; EPHRIN; 1.  
 SQ SEQUENCE 341 AA; 37849 MW; CB922F20B0D93E94 CRC64;

Query Match 33.3%; Score 616; DB 13; Length 341;  
 Best Local Similarity 41.8%; Pred. No. 1.3e-39;  
 Matches 141; Conservative 49; Mismatches 109; Indels 38; Gaps 11;

QY 24 LVSGISLEPYVNSANKRFQAEQGYLYLPQIGRLDLCPRAPPSPHSSPNYEFYKYL 83  
 DB 23 LPAKSLSEVWNSSQNPFXVGKGLVIVPEIGDKLDIICPKV---GDMGRD-YEFYKYL 77  
 QY 84 VAGAQRRCAPAPNLLTCDPDLRTTFQGYSPNLMGHERSHDYLIATSDG 143  
 DB 78 VKQDAECSCTILDPNVLTCKNPKEDTKFTKQFSPNLMGLEFQKFTVYITTSNG 137  
 QY 144 TREGLSLQGVCLTRGMKVLIVGQSPRG---GAVPRKPVSEMPERDRGAHSLSP 198  
 DB 138 TQGLLENNEGVCSTRSMKLMKVGDQPNADPDLPLDPRPYNEIKDPTSPSRKTER 197  
 QY 199 GKEN-----LQDPTSNATSR--GAEGLPPSPMPAVAGAAG-GIALLLVAGAGG 247  
 DB 198 GRENEVDNGSRMKGKTRQNNNSPGSVGIF--GSXPALFPAAGACVFLIILIV 255  
 QY 248 AMCWRRRAKPSRSRHPGSGFRGSL---GLGGGGMGPREAREGELIALRGGA 303  
 DB 256 LLKLRKRKR---KHSQ---RGTLSTLATRGAAQAGSEPSDIIIFLR---TT 304  
 QY 304 DPFQHYEYSGDIYGVYIVDPPQSPSPNITYKV 340  
 DB 305 ENNYCPYKESGDIYGVYIVDPPQSPSPNITYKV 341

RESULT 6  
 ID Q90232 PRELIMINARY; PRT; 334 AA.  
 AC Q90232;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
 DE Ephrin B2b.  
 OS Brachydanio rerio (Zebrafish) (Zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OC NCBI\_TaxID=7955;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=21290827; PubMed=11397014;  
 RA Chan J., Mably J.D., Serluca F.C., Chen J.N., Goldstein N.B.,  
 RA Thomas M.C., Cleary J.A., Brennan C., Fishman M.C., Roberts T.M.;  
 RT "Morphogenesis of prechordal plate and notochord requires intact  
 RT eph/ephrin b signaling."  
 RL Dev. Biol. 234:470-482(2001).  
 RL EMBL; AF375226; AAK64276.1; -  
 DR InterPro; IPR001799; Ephrin.  
 DR Pfam; PF00812; Ephrin.1.  
 DR ProDom; PD002533; Ephrin.1.  
 SQ SEQUENCE 334 AA; 36998 MW; 341497E0F83472BD CRC64;

Query Match 32.7%; Score 605.5; DB 13; Length 334;  
 Best Local Similarity 41.0%; Pred. No. 7.9e-39;  
 Matches 134; Conservative 45; Mismatches 113; Indels 35; Gaps 8;

QY 30 LEPYVNSANKRFQAEQGYLYLPQIGRLDLCPRAPPSPHSSPNYEFYKYLIVGAAG 89  
 DB 27 LESTVYNTSNKFPVPGQGLVLYPGIDKMDIVCPRIK-POSTGQTNLEHYRVIVPEQL 85  
 QY 90 RRCAPAPNLLTCDPDLRTTFQGYSPNLMGHERSHDYLIATSDGTREGLE 149  
 DB 86 ETCVHTKSMILLNCDPDDQVKTTFQFSPNLMGLEFQKGDVYIISTNSSTFGLD 145  
 QY 150 SLQGVCLTRGMKVLIVGQSPRGAV---PKPVSEMPERDRGAHSLPEQKENVLP 205  
 DB 146 NHGGVGRSKSMKVLIVAGQSPDTSFSAKHPTRNPKYIENKQDNTF---SKENDVS 200  
 QY 206 --DPTSNATRGAEGLPPSPMPAVAGAGLALLLVAGAGAGCW-----RR 253  
 DB 201 QIDMQNGESGSGXSG-----ESVSGAGSDVALFAGV--AGGAVIFLIITVALH 250

QY 254 REAKESSEHPPGSGFRGSGSLGCGGGMGPBAPGELGALRGGAADPPFCPEYK 313  
 DB 251 RHQHGHSAQCCSQGLPLNTLPKRGSGASGSGNNNGSEPSDIFPIRTSGM---YCPHYEK 307  
 QY 314 VSGDYGHPPYIVQDGPPOSPNNIYKV 340  
 DB 308 VSGDYGHPPYIVQEMPONPANITYKV 334

## RESULT 7

Q9W6H9 PRELIMINARY; PRT; 205 AA.  
 AC Q9W6H9;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Ephrin-B2 (Fragment).  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
 OC Xenopodinae; Xenopus.  
 NCBI\_TaxID=83355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97411149; PubMed=9259557;  
 RA Smith A., Robinson V., Patel K., Wilkinson D.G.,  
 RT "The EphA4 and EphA1 receptor tyrosine kinases and ephrin-B2 ligand  
 RT regulate targeted migration of branchial neural crest cells.",  
 RL Curr. Biol. 7:561-570(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20093673;  
 RA Hebling F.M., Saulnier D.M., Robinson V., Christiansen J.H.,  
 RA Wilkinson D.G., Brandt A.W.,  
 RT "Comparative analysis of embryonic gene expression defines potential  
 RT interaction sites for Xenopus EphA4 receptors with ephrin-B ligands",  
 RL Dev. Dyn. 216:361-373(1999).  
 DR EMBL; AF128844; AAD32610.1;  
 DR InterPro; IPR001799; Ephrin.  
 DR Pfam; PF00812; Ephrin; 1.  
 DR ProDom; PD002533; Ephrin; 1.  
 FT NON TER 1  
 SQ SEQUENCE 205 AA; 22256 MM; 7DEDD3427726C87 CRC64;

Query Match 18.4%; Score 340; DB 13; Length 205;  
 Best Local Similarity 37.0%; Pred. No. 1.2e-18;

Matches 84; Conservative 35; Mismatches 72; Indels 36; Gaps 7;

QY 128 EFRSHDYIIATSDGTREGLLESLQGVCLTRGMKVLRLVQSGP-----RGAVPRKPV 182  
 DB 1 EFGRRKDYIIISTNGSLEGVNCGGVCYTKAMKILMKVQDPFHHNRGASSTRBDH 60  
 QY 133 EYPM--ENDRGAAHSLFEGKENLPEDPTSNATSGAESPPLPPSPMPAVAGAAGLALLL 240  
 DB 61 ESGTNGKSTTSPHYNGEGSSTEGKNAGHSILGSEVAL-----FAGIASGSIIFIV 113  
 QY 241 GVAAGAGMCMRRRAKPSSEHPPGSGFRGSGSLGCGGGMGPBAPGEL 293  
 DB 114 IITVLTLTKRRHRKHSPOHT-----TSLSLTATPKRSGNNNG---SEPSDI 161  
 QY 234 GIALRGGAADPPFCPEYKXVSGDYGHPPYIVQDGPPOSPNNIYKV 340  
 DB 162 IIPLR---TAGGVFCPEYKXVSGDYGHPPYIVQEMPONPANITYKV 205

## RESULT 8

Q9U3M2 PRELIMINARY; PRT; 237 AA.  
 AC Q9U3M2;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE C43F9.8 protein.

GN C43F9.8.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Felodierinae; Caenorhabditis.  
 NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Mortimore B.J.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;

QY 11 VRVGAALLGLVGLVS-GSLPEVYWNANKRFOAG-GVLYYPOIGRLDLCPRRAP 68  
 DB 1 MQATATLTLSPFPGMAKPIDIMWISNPFDVSNMDHVSVHIGDVSIRCKSDPT 60  
 QY 69 GPHSSPNYEFYLYLVGAGQRCAPAPNLLTCDDPDDLTFTYFQESPLKHE 128  
 DB 61 G-----KYYSYTYWVSDEYDHCFL-SKPRLVGACDQTNASINIVFRSSTPPGGE 114  
 QY 129 FRESHDYIIA-----TSDGTREGLLESLQGVCLTRGMKVLRLVQSG 169  
 DB 115 FQGRKDYIIISTNGSLEGVNCGGVCYTKAMKILMKVQDPFHHNRGASSTRBDH 174  
 QY 170 SPRGAVPRKPVSEMERDGAHS 195  
 DB 175 DRRGIENPK-FAARTLKDRDAHS 198

Query Match 11.6%; Score 214.5; DB 5; Length 237;  
 Best Local Similarity 25.7%; Pred. No. 6.1e-09;

Matches 53; Conservative 41; Mismatches 83; Indels 29; Gaps 6;

QY 09U474 PRELIMINARY; PRT; 279 AA.  
 AC Q9U474;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE VAB-2 (Hypothetical protein Y37E11AR.6).  
 GN VAB-2 OR Y37E11AR.6.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Felodierinae; Caenorhabditis.  
 NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=N2;  
 RX MEDLINE=20084449; PubMed=10619431;  
 RA Chin-Sang I.D., George S.E., Ding M., Moseley S.L., Lynch A.S.,  
 RA Chisholm A.D.;  
 RT "The ephrin VAB-2/EFN-1 functions in neuronal signaling to regulate  
 RT epidermal morphogenesis in C. elegans",  
 RL Cell 99:761-780(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.",  
 RL Science 282:2012-2018(1998).





```

DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE Ephrin-A6 (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauiria; Aves; Neognathae; Galliformes; Psittacidae; Psittacinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Menzel P., Valencia F., Godement P., Dodelat V.C., Pasquale E.B.;
RT "Ephrin-A6, a new ligand for EphA receptors in the developing visual
system.";
RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF317286; AAK0944.1; -
DR InterPro; IPR001799; Ephrin.
DR Pfam; PF00812; Ephrin; 1.
DR PRINTS; PR01347; Ephrin; 1.
DR ProDom; PD002533; Ephrin; 1.
DR PROSITE; PS01299; EPHRIN; 1.
DR NON TER 1
SQ SEQUENCE 202 AA; 22624 MW; 91E2716FP9135F9 CRC64;

Query Match 9.6%; Score 178.5; DB 13; Length 202;
Best Local Similarity 27.5%; Pred. No. 3e-06;
Matches 56; Conservative 27; Mismatches 84; Indels 37; Gaps 7;

QY 33 VYWNANKRFGQEGVYVYPOIGDRDLCPARPSPGSHSPNYFYLVGAGQGRG 92
DB 25 VYWNNSNRF-LQDDYSLOVSNIDHDIYCHPYSAPTPMA-----ESFLFWDEGIRGC 79
QY 93 EAPPAPLLITCDR--PDLDRTIKFOEYSPNLMGHFESHHDYLIAT-SDGTBGL 148
DB 80 SETPAFKMECNKPPAPFVPVPRFEKIQRFPSLGFEPFGETYYIISVPTGSGS---- 135
QY 149 ESLGGVCLTRGKMTLVKRGSPRGAVPRKVSSEMPERDGAHSLPEKENVPGDPT 208
DB 136 ---AGRLCKRVSYCCR-----ASTPEFLTEVPNSQRRG-----GGPE 171
QY 209 SNATRGAGEGPLPPSPMAVAGAA 232
DB 172 GDGSPRDAAP-PQRSRTLVALA 195

RESULT 12
Q9D7K8 PRELIMINARY; PRT; 205 AA.
AC Q9D7K8;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Adult male tongue cDNA, RIKEN full-length enriched library,
DE clone:331004015, full insert sequence.
GN EFNA1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
RA Saito T., Okazaki Y., Gojopori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Iono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaudo M.F.,

```

```

RA Brownstein M.J., Bulb C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima U., Mazzarelli U., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK009144; BAB26102.1; -
DR MGD; MGI:103236; Etnal.
DR InterPro; IPR001799; Ephrin.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00812; Ephrin; 1.
DR PRINTS; PR01347; Ephrin; 1.
DR ProDom; PD002533; Ephrin; 1.
DR PROSITE; PS01299; EPHRIN; 1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN 1.
DR SEQUENCE 205 AA; 23772 MW; E37E55767459A4BC CRC64;

Query Match 9.0%; Score 167; DB 11; Length 205;
Best Local Similarity 26.1%; Pred. No. 2.3e-05;
Matches 43; Conservative 34; Mismatches 76; Indels 12; Gaps 4;

QY 18 LILGVLVGLSLSPYWNANKRFGQEGVYVYPOIGDRDLCPARPSPGSHSPNYE 77
DB 8 LILGCSLAADRHIVFNMSNPKFREE-DYTHVQLNDYLDICHPYDSDSV-ADAAME 65
QY 78 FYLVYVGAQGRGRCAPAPPLLITCDRPL--DLRTIKFOEYSPNLMGHFESHHD 134
DB 66 RYTLVNEHOEYVACOPQSKDOVRNKNRPSAKHGPEKLSVFORPTPLIGKEFEGHS 125
QY 135 YXLIATSDGTREGLESLGGVCLTRGKMTLVKRGSPRGAVPRK 179
DB 126 YXYSKPIYHOE-----SQCCLKVTVNGKITHNPQAHVPOE 163

RESULT 13
Q9CZS8 PRELIMINARY; PRT; 206 AA.
AC Q9CZS8;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 10 days embryo cDNA, RIKEN full-length enriched library,
DE clone:2610529M21, full insert sequence.
GN EFNA4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
RA Saito T., Okazaki Y., Gojopori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Iono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaudo M.F.,
RA Brownstein M.J., Bulb C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima U., Mazzarelli U., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuki S.,

```

RA Hayashiraki Y.:  
 RT "functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL; AK012195; BAB28092.1; -  
 DR MGI; 106643; Efn4.  
 DR InterPro; IPR001799; Ephrin.  
 DR Pfam; PF00812; Ephrin.1.  
 DR PRINTS; PR01347; EPHRIN.  
 DR ProDom; PD002533; Ephrin.1.  
 DR PROSITE; PS01299; EPHRIN.1.  
 DR SEQUENCE 206 AA; 22859 MW; 675E32971D1C6EBC CRC64;

Query Match 8.8%; Score 163.5; DB 11; Length 206;  
 Best Local Similarity 28.1%; Pred. No. 4.3e-05;  
 Matches 61; Conservative 17; Mismatches 80; Indels 59; Gaps 10;

QY 32 PYVNSANKRFOAGGYLYPQIGRLDILCPRARPPGPHSSPNVEFYKLYLVGAQGR 91  
 DB 29 PIYVNSNPRL-LRGDAVVELGFNDYLDIFCPHYESPGE--ETFAIYIVDSGYEA 85  
 QY 92 CEAPPAIPL-LITCDRPPD---DLRTIKFOEYSPNMGHERSHHDYIITATSDGTREG 147  
 DB 86 CIEGANSFORANCSMPFAPFSPVSEKIQRTPLPLGFELPGETYYIISVPTPESPG 145  
 QY 148 -LESLOGVCLTRGMKYLRLVGSQSPGAVPRKPVSEMERDRGAH-SLEPKENLPG 205  
 DB 146 RCLRLGVSVCC-----KESGSHSAMP-----VG 170  
 QY 206 DPTSNATS--RGAEGLPPSPMPAVAGAGLALLL 240  
 DB 171 SPGESGTSGRGSHAPSP-----LCILLL 194

## RESULT 14

Q9N178 PRELIMINARY; PRT; 675 AA.

AC Q9N178; 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Type X collagen.  
 CN COL10A1.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.  
 NCBI\_Taxid=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21015405; PubMed=11130976;  
 RA Nielsen V.H., Bendixen C., Arndbjerg J., Sorensen C.M., Jensen H.E.,  
 RA Shukri N.M., Thomsen B.;  
 RT "Abnormal growth plate function in pigs carrying a dominant mutation  
 in type X collagen."  
 RT Mamm. Genome 11:1087-1092(2000).  
 RL EMBL; AF22861; AAF37271.1; -  
 DR InterPro; IPR001073; C1Q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1Q.1.  
 DR Pfam; PF01391; Collagen.8.  
 DR PRINTS; PR00007; COMPLEMENTC1Q.  
 DR ProDom; PD000007; COLLAGEN.2.  
 DR SMART; SM00110; C1Q.1.  
 DR PROSITE; PS01113; C1Q.1.  
 DR Collagen.  
 SQ SEQUENCE 675 AA; 65447 MW; 26397B10310363F9 CRC64;

Query Match 8.4%; Score 156; DB 6; Length 675;  
 Best Local Similarity 28.6%; Pred. No. 0.00063;  
 Matches 106; Conservative 17; Mismatches 118; Indels 130; Gaps 26;  
 QY 1 MGPFHSGPGGVRYGALLLGLVGLSLPYYNSANKRFOAGGYLYPQI-GDR-L 58  
 DB 212 MGPP--GPPGV-----GKR--GENGFPGGIGKDRGF 240

QY 59 DLICPRARPPGPHSSPNVEFYKLYLVGAQGR-CEAPPAIPLITCDRPPDLRTIKF 117  
 DB 241 PGESGPAIPLPPGPP-----GEGRBEIGKPAFG---AAGQGL----- 278  
 QY 118 QEYSPNMGHERSHHDYIITATSDGT---RGLSELQGVCLTRGMKYLRLVGSQSPG 173  
 DB 279 ---FGTKGHPGAG---MAGPPGAPGPKGGLPKG---QRG-----PIG 315  
 QY 174 GAVPRKPVSEMERDRG-AAHSLPEKENVLPDPTSNATSRGAEGLPPSPVAVGAA 232  
 DB 316 --LPGAFGA---KGEQSPAGHPPEPDLTGPFG-----SRGQSPKGLPANNVGPBK 362  
 QY 233 GGLALLLVAGAGAGAMCWRRR---AKPSESHPG-----PGSTGRGSLGLGGGG 282  
 DB 363 GEIG--LAGPAGPPGAGKGERGPGSLDGKPGVPGEPGINKGNPGLPGPKDDPQIGPPG 420  
 QY 283 M---GPREA--PGEIGL-LRG-----GAADPFCEHYEKVSGDYHPYIVQD 328  
 DB 421 LPGVGPAGAGVGHNGEAGPRGAPGIPGTRGIPGPIPGFPGSKXDPGNP-----G 474  
 QY 329 PPGSPNIYK 339  
 DB 475 PP-GPAGIATK 484

## RESULT 15

Q93431 PRELIMINARY; PRT; 229 AA.

AC Q93431; 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Ephrin A-11.  
 OS Brachydanio rerio (Zebrafish) (Zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 NCBI\_Taxid=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Durbin L., Brennan C.H., Shiomi K., Cooke J.;  
 RT "Eph signalling is required for segmentation and differentiation of  
 the somites."  
 RT Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; A006838; CA07264.1; -  
 DR InterPro; IPR001799; Ephrin.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR Pfam; PF00812; EPHRIN.1.  
 DR PRINTS; PR01347; EPHRIN.  
 DR ProDom; PD002533; Ephrin.1.  
 DR PROSITE; PS01299; EPHRIN.1.  
 DR PROSITE; PS00290; IG\_MHC; UNKNOWN 1.  
 SQ SEQUENCE 229 AA; 26115 MW; 8684462F67A6F5C CRC64;

Query Match 8.4%; Score 155; DB 13; Length 229;  
 Best Local Similarity 27.2%; Pred. No. 0.00022;  
 Matches 56; Conservative 23; Mismatches 93; Indels 34; Gaps 7;

QY 33 PYVNSANKRFOAGGYLYPQIGRLDILCPRARPPGPHSSPNVEFYKLYLVGAQGR 92  
 DB 25 PYVNSNPRL-LRGDAVVELGFNDYLDIFCPHYESPGE--ETFAIYIVDSGYEA 85  
 QY 93 CEAPPAIPL-LITCDRPPD---DLRTIKFOEYSPNMGHERSHHDYIITATSDGTREG 147  
 DB 86 CIEGANSFORANCSMPFAPFSPVSEKIQRTPLPLGFELPGETYYIISVPTPESPG 145  
 QY 150 SL--QGVCITRGMKYLRLVGSQSPGAVPRKPVSEMERDRGAH-SLEPKENLPG 207  
 DB 134 PLHHGQECRLKVDV-----GPHGSKKKKQKVEEETEGMAAGVANNPNNRLPAD 188  
 QY 208 TSNATSRGAEGLPPSPMPAVAGAG 233

Db 189 -----PIAMTPVQBSVG 201

# RESULT 16

Q9ESQ2 PRELIMINARY; PRT; 1691 AA.

AC Q9ESQ2;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Type IV collagen alpha 5 chain.

GN COL4A5.

OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]  
RX MEDLINE=20536494; PubMed=10965041;

RA Saito K., Naito I., Seki T., Ohashi T., Kimura E., Momota R.,

RT "Differential Expression of Mouse a5(IV) and a6(IV) Collagen Genes in

Epithelial Basement Membranes."

RL J. Biochem. 128:427-434(2000).

DR EMBL; AB041350; BAB13673.1; -

DR MGD; MGI:88456; Col4a5.

DR InterPro; IPRO00087; Collagen.

DR Pfam; PF01413; C4; 2.

DR Pfam; PF01391; Collagen; 24.

DR ProDom; PD000007; Collagen; 3.

DR ProDom; PD003923; Procollagnc4; 1.

DR SMART; SM00111; C4; 2.

DR Collagen.

DR SMAR; SM00111; C4; 2.

DR SMAR; SM00111; C4; 2.

DR SMAR; SM00111; C4; 2.

DR SMAR; SM00111; C4; 2.

DR SMAR; SM00111; C4; 2.

DR SMAR; SM00111; C4; 2.

DR SMAR; SM00111; C4; 2.

DR SMAR; SM00111; C4; 2.

DR SMAR; SM00111; C4; 2.

DR SMAR; SM00111; C4; 2.

DR SMAR; SM00111; C4; 2.

DR SMAR; SM00111; C4; 2.

DR SMAR; SM00111; C4; 2.

DR SMAR; SM00111; C4; 2.

DR SMAR; SM00111; C4; 2.

DR SMAR; SM00111; C4; 2.

DR SMAR; SM00111; C4; 2.

DR SMAR; SM00111; C4; 2.

DR SMAR; SM00111; C4; 2.

GN COL1A1.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

OC Xenopodinae; Xenopus.

OX NCBI\_TaxID=8355;

RN [1]  
RX MEDLINE=20536494; PubMed=10965041;

RA Saito K., Naito I., Seki T., Ohashi T., Kimura E., Momota R.,

RT "Differential Expression of Mouse a5(IV) and a6(IV) Collagen Genes in

Epithelial Basement Membranes."

RL J. Biochem. 128:427-434(2000).

DR EMBL; AB041350; BAB13673.1; -

DR MGD; MGI:88456; Col4a5.

DR InterPro; IPRO00087; Collagen.

DR Pfam; PF01413; C4; 2.

DR Pfam; PF01391; Collagen; 24.

DR ProDom; PD000007; Collagen; 3.

DR ProDom; PD003923; Procollagnc4; 1.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

GN COL1A1.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

OC Xenopodinae; Xenopus.

OX NCBI\_TaxID=8355;

RN [1]  
RX MEDLINE=20536494; PubMed=10965041;

RA Saito K., Naito I., Seki T., Ohashi T., Kimura E., Momota R.,

RT "Differential Expression of Mouse a5(IV) and a6(IV) Collagen Genes in

Epithelial Basement Membranes."

RL J. Biochem. 128:427-434(2000).

DR EMBL; AB041350; BAB13673.1; -

DR MGD; MGI:88456; Col4a5.

DR InterPro; IPRO00087; Collagen.

DR Pfam; PF01413; C4; 2.

DR Pfam; PF01391; Collagen; 24.

DR ProDom; PD000007; Collagen; 3.

DR ProDom; PD003923; Procollagnc4; 1.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

DR SMART; SM00111; C4; 2.

RA None;  
 RT "genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.",  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA pauley A., Gattung S.;  
 RT "The sequence of C. elegans coemid T15B7.",  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Waterston R.;  
 RT "Direct Submission.",  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF022885; AAB69961.1; -  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR002486; Col cuticle\_N.  
 DR Pfam; PF01391; Collagen\_3.  
 DR Pfam; PF01484; Col cuticle\_N; 1.  
 DR Hypothetical protein.  
 KW SEQUENCE 325 AA; 31591 MW; 6538D4537C9EC202 CRC64;  
 SQ  
 Query Match 8.1%; Score 149; DB 5; Length 325;  
 Best Local Similarity 26.0%; Pred. No. 0.00095;  
 Matches 92; Conservative 25; Mismatches 109; Indels 128; Gaps 23;  
 QY 13 VGALLILGVLVSGSL-----EPVYNSANKRFQAEQ---YLYPQIGDRLLDLIC 62  
 DB 1 MSASTLVTAASAAGIAIVCVFTVMGFINDINSFDEKIGEKKEQEQL--WQAMI 58  
 QY 63 PRARPPPHSPHYEYKYLVC-----GAGRRCFA-PRAPNLLTCDRLDL 111  
 DB 59 PTTSPSSGSS-----FLLRNKRQAQCNGASRGCPAPRPP 96  
 QY 112 RFTIKPQEVSPNMGHFRSHNDYIIATSDTREGLESLOG--GYCLTRGMKVLRLVQ 169  
 DB 97 -----PGQPGAPGSGHNG-----LAQPGSGARINPATGPGFCIT----- 133  
 QY 170 SPRGAVPRKPVSEMPMERDRGAHSLERKENTLPDPTSNATSRAGPLPPSPMPAVA 229  
 DB 134 CPAGAPGAPAP-----PGA-----PGPKNNQPGAPAPAS--GGRAP-PGPRGP-A 175  
 QY 230 GAAGGLALLLLGVAGAGACWRRRAKPSRSRHPG-PSFGSGG--SLGLGGGGGM--- 283  
 DB 176 GDAG-----SPGQPGHFGSPGPNPGSGGSRSLPLPPSPSRPP 212  
 QY 284 -GPREAPGELGIALRGGAADP-PFCPHYEKVSGDYGHFVYIVDGPQSPPN 335  
 DB 213 PGF-AGGPGQPG--HSGGAGSPGPGP-----PGSPGPGHSGNDGVGAPGN 257  
 RESULT 19  
 Q17208 PRELIMINARY; PRT; 569 AA.  
 ID Q17208;  
 AC Q17208;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Collagen (Fragment).  
 OS Bombyx mori (Silk moth).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
 OC Bombycoidea; Bombycidae; Bombyx.  
 OX NCBI\_TaxId=7091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BURROPEAN 200X300; TISSUE=IMAGINAL WING DISC;  
 RA MEDLINE=97148965; Pubmed=8995790;  
 RX Charayre P.P., Besson M.M., Fouchet J.J., Bosquet G.G.;  
 RT "Identification of a Bombyx collagenous protein with multiple short  
 RT domains of Gly-Xaa-Yaa repeats: cDNA characterization and regulation

RT of expression".  
 RL Insect Biochem. Mol. Biol. 26:677-685(1996).  
 DR EMBL; Z30348; CA83002.1; -  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001419; Glutenn.  
 DR Pfam; PF01391; Collagen; 5.  
 DR PRINTS; PR00210; GLUTENIN.  
 FT NON\_TER 1  
 SQ SEQUENCE 569 AA; 56583 MW; FE371482F0A3B0AF CRC64;  
 Query Match 8.1%; Score 149; DB 5; Length 569;  
 Best Local Similarity 29.3%; Pred. No. 0.0018;  
 Matches 51; Conservative 14; Mismatches 77; Indels 32; Gaps 6;  
 QY 171 PRGAVPRKPVSEMPMERDRGAHSLERKENTLPDPTSNATSRAGPLPPSPMPAVA 230  
 DB 1 PGQGGQPKKPGQ-----PGYPGQPGQPGYPGQ--GQPGAPGQPGQPGQPGGTPG 50  
 QY 231 AAGGLALLLLGVAGAGACWRRRAKPSRSRHPG-----PGSPGSGSLGGGGGM 283  
 DB 51 QAG-----QPTYPQGG-----QPIKPAQGHGPGQPGQPGQPGQPGQPGQ 99  
 QY 284 GPREAPGELGIALRG-----GAADPFCPHYEKVSGDYGHFVYIVDGPQSP 335  
 DB 100 PGQPGQPGQPGQPGQPGQPGQPGQPGQPGQPGQPGQPGQPGQPGQPGQPGQ 153  
 RESULT 20  
 Q19111 PRELIMINARY; PRT; 316 AA.  
 ID Q19111;  
 AC Q19111;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE F02D10.1 protein.  
 GN F02D10.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxId=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Swindburne U.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; Pubmed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C. elegans: A platform for  
 RT investigating biology.",  
 RL Science 282:2012-2018(1998).  
 DR EMBL; Z67990; CA91932.1; -  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR002486; Col cuticle\_N.  
 DR Pfam; PF01391; Collagen; 2.  
 DR Pfam; PF01484; Col cuticle\_N; 1.  
 DR Pfam; PF01484; Col cuticle\_N; 1.  
 SQ SEQUENCE 316 AA; 29444 MW; 4D76D5BA07923499 CRC64;  
 Query Match 8.0%; Score 148.5; DB 5; Length 316;  
 Best Local Similarity 28.5%; Pred. No. 0.001;  
 Matches 49; Conservative 12; Mismatches 78; Indels 33; Gaps 6;  
 QY 168 GQSPRGAVPRKPVSEMPMERDRGAHSLERKENTLPDPTSNATSRAG 216  
 DB 144 GASGKGAAAPCEPSTPPCQ-PCPAGPDPGPGCTGEGGPGAPASPAAGSPGAPG 202  
 QY 217 EGPLPPSPMPAVAGAGGLALLLLGVAGAGACWRRRAKPSRSRHPG-GSPRGGS 275  
 DB 203 PGAPGPGNDGQPGQPGGPG--QDGAASAG-----EAGPGAGPGPAGPAGP 253  
 QY 276 GLGGGGGMP-----EAPGELGIALRGGAADPFCPHYEKVSG 316  
 DB 254 GSGAGGPGPGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 305

```

RESULT 21
ID Q61434 PRELIMINARY; PRT; 1140 AA.
AC Q61434;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Collagen (Fragment).
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Abe N., Muragaki Y., Yoshitaka H., Inoue H., Nimomiya Y.;
RT "Identification of a novel collagen chain represented by extensive
RT interruptions in the triple-helical region."
RL Cell. Mol. Biol. Res. 196;576-582(1993).
DR EMBL; D17546; BA04483.1; -.
DR HSSP; P39061; IKOE.
DR MGD; MGI:88449; Col15a1.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 8.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Collagen; 1.
FT NON_TER 1
SQ SEQUENCE 1140 AA; 115156 MW; 880C7B6862B3BDFE CRC64;

Query Match 7.9%; Score 146.5; DB 11; Length 1140;
Best Local Similarity 23.7%; Pred. No. 0.0062;
Matches 92; Conservative 22; Mismatches 107; Indels 167; Gaps 21;

QY 2 GPPHSGPGGVAVGALLIGVIG-LVSGLSLEPYVWNSANKRFQABGGVLYPQIGRDL 60
DB 164 GPP--GPAQPQ-----GPAQPVQSPNSQPV----- 187
QY 61 LCPRAR-PPGPPSPNVEFYKLYLVGAQGRCE----- 93
DB 168 --PGAQGPFGQGPGRKQ-----GTPGRDGEFGDGEGRGDDGPGQPGTDPDVG 237
QY 94 -----APPAENLLTCDRPDLRFTIKQEVSPNLMGHEFRSHDYIIAT 140
DB 238 PKGKGDGPGIGRGPGRGP-----PGPPGPFGRQDKLTIDME 274
QY 141 SDGTREGLESLOGVCILTRGKAVLIRVQSGAVPRKP--VSEMPMERD--GAASLE 197
DB 275 GSGFGSDIESLRG-----PRGPPGPPGPGVGLPGEPRFQINSYA 317
QY 198 FKEKMLPDPTSNATSRGAEGP--LP-PPSMPAVAGAGLALLLVAGAGAMCWRRR 254
DB 318 PGPALPLPVP-----GKEGPPRGPGRPPGPKGKP-----PGVAGKSGV----- 360
QY 255 RAKPESRHPGP-GSFGGSGIGLGGGGMGKPREAPPELIGALRG--GAADPFQPHY 311
DB 361 ---GDVGI-PGPKSGKGLDIPGMPKSGLA---GSPGPVGPGRPPGPPGPGFAAGF 413
QY 312 EKVSQGYGHPVIVQ-----DGPQSP 333
DB 414 DMEGSG-GIFLWTTARSSDGLQGPPGSP 440

RESULT 22
ID Q91718 PRELIMINARY; PRT; 1491 AA.
AC Q91718;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Alpha-1 type II collagen.

```

```

OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxId=8335;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92011898; PubMed=1918153;
RA Su M.W., Suzuki H.R., Bleker J.C., Solursh M., Ramirez F.;
RT "Expression of two nonallelic type II procollagen genes during
RT laevis embryogenesis is characterized by stage-specific production of
RT alternatively spliced transcripts."
RL J. Cell Biol. 115:565-575(1991).
DR EMBL; M63586; AA49679.1; -.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; WVF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; WVC; 1.
DR ProDom; PD000007; Collagen; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; WVC; 1.
DR PROSITE; PS01208; WVFC; 1.
KW Collagen.
SQ SEQUENCE 1491 AA; 142495 MW; 43026FF08FB0314 CRC64;

Query Match 7.9%; Score 146.5; DB 13; Length 1491;
Best Local Similarity 31.1%; Pred. No. 0.0085;
Matches 56; Conservative 11; Mismatches 80; Indels 33; Gaps 8;

QY 171 PGCAVPRKPVSEMPMERDRG-AAHSLPEKENTLPDPTSNATSRGAAGPLPPSMPAVA 229
DB 123 PKGPPGPGSGEGRGRGDXGTGAPRGRGEP-----GTGNGPAGPFPGP 178
QY 230 GAAGLALLIG--VAGAGAMCWRRR-----RAKPSERHGPQSF--GRGSLG 278
DB 179 GIGGPFPAAGMTGTFEKKAGAGMGVAGQPMGPRGPPGPGAGAGPQGNPPEPS 238
QY 279 GGGG-MGPR-----EAEPDELIGALRG-----GAADPFQPHYEVSQGYHP 321
DB 239 GAGGPMGPPGPGPSGKFGDDSEAKRGKSGRGPFGPGAGRGFTGTGLPQVKHRYTP 298

RESULT 23
ID Q62001 PRELIMINARY; PRT; 1774 AA.
AC Q62001; Q60672;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Procollagen, type XVIII, alpha 1 precursor (XVIII) collagen
DE (Procollagen, type XVIII, alpha 1) (Alpha-1 type XVIII collagen).
GN COL18A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=PRIMARY TAIL CULTURE;
RX MEDLINE=94245707; PubMed=8188673;
RA Rehn M., Hintikka E., Pihlajaniemi T.;
RT "Primary structure of the alpha 1 chain of mouse type XVIII collagen,
RT partial structure of the corresponding gene, and comparison of the
RT alpha 1 (XVIII) chain with its homologue, the alpha 1(XV) collagen
RT chain."
RL J. Biol. Chem. 269:13929-13935(1994).
RN [2]
RP SEQUENCE OF 1-562 FROM N.A.
RX MEDLINE=94240112; PubMed=8183894;
RA Rehn M., Pihlajaniemi T.;

```



```

RL Genomics 61:113-124(1999).
DR EMBL; AF169387; AAD50449.1; -.
DR MGI; MGI:104688; Col4a3.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001442; Procollagnc4.
DR InterPro; IPR000504; Rna_rec_mot.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 21.
DR ProDom; PD000007; Collagen; 6.
DR ProDom; PD003923; Procollagnc4; 1.
DR SMART; SMO0111; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
KW Collagen.
SQ SEQUENCE 1669 AA; 161769 MW; 30576E59739A47B2 CRC64;

Query Match
Best Local Similarity 24.3%; Pred. No. 0.011;
Matches 97; Conservative 34; Mismatches 134; Indels 134; Gaps 22;

QY 4 PHSQGVGVVALLLGLGL-----VSGLSLEP-----VYNSANKRFQAEQGY 48
DB 1008 PHSQGVGV-----MGIMGVPGPKGRKGTSLPLGLAPRGRTGTHGPQDK--GEPGY 1056
QY 49 VLTPQIGRLDLCPRAPPGPHSSPNYEFYKLYLVGAGQGRCE----- 93
DB 1057 SEGARP-----PPGPKDDP-----LPQDKKKGERGVPPGPGSGPACP 1097
QY 94 -AAPAPNLLITCDRP-----DLDRFTTKFOEYSPNIMGHEFRSHNDYIATSDGTREG 147
DB 1098 DGAPSPSPSPGHPGKFPAGDLGL-----KQCKGFPFGP-----STGPPGPPG 1141
QY 148 LESLQGGVCL-----TRGKVLRLVQSPRGAVPRKVSSEMERDRGAA 123
DB 1142 LPLGPGPKMGMDGQGRDIPPPGKGTGLGAYPGPKGS--PGVGA---KGRGV- 1194
QY 194 HSLBEGKENTLPDPTSNATSGAGCPPLPPSPAPVAGAGLALLLVAGAGGAMCWR 253
DB 1195 ----PGLSGLPG-----RKGVMDVPGQPPGAPGPPGALIPGPKDRLPLG 1243
QY 254 RRAKPSERHNG-PGSPRG-----GSLGLGG--GGMPPRAELEGIALRG 300
DB 1244 LRNGPGEPPGPPGPIGKIGKDGKFGNGPPGKGLPRTVGDWGP-PGFPAPGTPGLPG 1302
QY 301 GAADPPCFPHYKXVGDYGHFVY---IYQDGP--PQSP 334
DB 1303 VRGDPGF-PGFPPIKGEKNGPFLGPIGHPGVPGKPP 1340

RESULT 26
ID 017805 PRELIMINARY; PRT; 305 AA.
AC 017805;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Flaf7.1 protein.
GN Flaf7.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
CX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurtry A.A.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RW [2]
RW SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
KW none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z81503; CAB04111.1; -.

```

```

DR InterPro; IPR000087; Collagen.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01484; Col_cuticle_N; 1.
SQ SEQUENCE 305 AA; 29497 MW; 161BFA90D024C2E6 CRC64;

Query Match
Best Local Similarity 22.5%; Pred. No. 0.0016;
Matches 83; Conservative 21; Mismatches 122; Indels 143; Gaps 12;

QY 11 VEVGALLLLG-----VLGVSGLSLEPVYWN-----SANK----- 40
DB 5 VALKAYRLIAFAVAFCLTSLVSVCTILPMYNYNSGKRVLLVGVSPCKKSANAEVFEV 64
QY 41 -----RQABGGVLYTPQIGRLDLCPRAPPPGPHSSPNYEFYKLYLVG 86
DB 65 NHIRASATNSTSARHAGYGY-----AQPQGGGG-----GG 97
QY 87 AGRRCAPAPNLLITCDRPDLRFTTKFOEYSPNIMGHEFRSHNDYIATSDGTRE 146
DB 98 GQSTCCRPGRPGAGTGRKG-----RPGAPGAMG----- 128
QY 147 GLESLQGGVCLTRGKVLRLVQSPRGAVPRKVSSEMERDRGAHSLBPKENLPD 206
DB 129 -----MPGNPGKGGSGCHPVVPTPCPCPG--RPGPPGPPGR 165
QY 207 PHSN-----ATSRGA-----EGFLPPSPMAPVAGAGLALLLVAGAGGAMCW 251
DB 166 PGSDQGPAPGAPGAPAPPPGPPGKFPAGDNGSGRAGAPQPDNAHGYGGVGAAPG 225
QY 252 RRRAPKPSERHNGPSPRGSGSLGLGGGGV---GPREAPBELGIALRGGAADPPF 307
DB 226 AGPRAGPAGHPPGSGGGRPPAPGAPGQPPRGP-DGHPQPPGRPGSGSGGNV 284
QY 308 CPHYKXVS 316
DB 285 CPXYCAIDG 293

RESULT 27
ID 094620 PRELIMINARY; PRT; 308 AA.
AC 094620;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Cuticle preprocollagen.
GN COL-2.
OS Meloidogyme incognita (southern root-knot nematode).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchidae; Heteroderidae; Meloidogyminae; Meloidogyme.
CX NCBI_TaxID=6306;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-CUTICLE;
RX MEDLINE=98324411; PubMed=9662034;
RA Wang T., Deom C.M., Husey R.S.;
RT "Identification of a Meloidogyme incognita cuticle collagen gene and
RT characterization of the developmental expression of three collagen
RT genes in parasitic stages.";
RT Mol. Biochem. Parasitol. 93:131-134(1998).
DR EMBL; U68729; AAC48358.1;
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF01484; Col_cuticle_N; 1.
KW Collagen.
FT CHAIN
SQ SEQUENCE 308 AA; 29361 MW; 825AEF0249PFC78 CRC64;

```

```

Query Match
Best Local Similarity 7.9%; Score 145.5; DB 5; Length 308;
Matches 46; Conservative 14; Mismatches 67; Indels 35; Gaps 6;

```



```

QY 171 PRGAVPRKVSSEMPERDGAHSLPEKKNLPGDPTSNATSRGABPLPPSPPAVAG 230
DB 155 PCRGGGPGGPGPPGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPG 204
QY 231 AAGGALLLLGAVAGAGAMCRRRRRAKPESEHHPG---PGSPFGGSGISGL-GGGGGNGP 285
DB 205 QPGG-----PGQPGGA-PA-OSQEPGAPAPPGPPGAGAGPGGPGGNGPAGQPGPGGPGP 255
QY 286 R-----BAEPGELGIALRGAGADPPPCPHYKVS 316
DB 256 KPPPPGPGGPGGNDGSPGQPGPPGPGTNGSGEKGICPKYCAIDG 297

RESULT 28
Q9ND02
AC Q9DD02 PRELIMINARY; PRT; 680 AA.
DB 01-JUN-2001 (TREMBlrel. 17, Created)
DB 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DB 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DB Procollagen, type IX, alpha 1.
GN COL9A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO.
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai U., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Izawa A., Fukunishi Y., Komori H., Adachi U., Fukuda S.,
RA Aizawa K., Izawa A., Fukunishi Y., Komori H., Adachi U., Fukuda S.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissel C., King B., Kochava H.,
RA Kuenli P., Lewis S., Matsuo Y., Nakai T., Pesole G., Quackenbush J.,
RA Schmitt L.M., Stabili F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Futuro M., Aono H., Baldarelli R., Barish G.,
RA Blake U., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamaly M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima U., Mazzaeili U., Mombereis P.,
RA Norone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohetsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DB EMBL; AK011547; BAB27690.1; -.
DB MGI; MGI:88465; Col9a1.
DB InterPro; IPR000087; Collagen.
DB Pfam; PF01391; COL1F; 1.
DB PRINTS; PR01582; KV3CHANNEL.
DB ProDom; PD000007; Collagen; 3.
KM Collagen.
SQ SEQUENCE 680 AA; 64733 MW; 5C0FB3E9992B4D91 CRC64;

Query Match 7.8%; Score 145; DB 11; Length 680;
Best Local Similarity 24.0%; Pred. No. 0.0045;
Matches 87; Conservative 22; Mismatches 133; Indels 122; Gaps 17;

```

```

DB 200 -----PRGQHKKEBPDQGLGEVGDQPPGQGLRGITGIVGDKXKRGARGFDEPGQ 254
QY 175 AVP-----RKVSSEMPERDGAHSLPEKKNLPGDPTSNATSRGABPLPPSP 224
DB 255 GIPGAGDQGGPGPPGPGTGTGEGDRGI-----QSGRGIPGSPGPGXGDTGLGVGDGRDIFG 309
QY 225 MPVAGAG-----GLALL--LLGAVAGAGAMCRRRRRAKPESE-----SRHPG--- 265
DB 310 MPGRKGAAGKRGPPGPDVGLGGLPGVPGIPGAKVYAGKMYGAPGKPGQGLSGSGKPGQG 369
QY 266 -PGSFGGSGISGL-GGGGGNGPRAE--PGELGIALRG--GAADPPCPHYKVSVDYG 319
DB 370 PPGVGPGRGPRGLPGSGRGVPGEGSPGIPGKLGSGVSGPLGPGPGLGPKMGKDRGVG 429
QY 320 HP 321
DB 430 EP 431

RESULT 29
Q9NZ06
AC Q9NZ06 PRELIMINARY; PRT; 1745 AA.
DB 01-OCT-2000 (TREMBlrel. 15, Created)
DB 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DB 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DB Collagen type V alpha 3 chain.
GN COL5A3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA; AND HEART;
RX MEDLINE=20187594; PubMed=10722718;
RA Imanura Y., Scott I.C., Greenspan D.S.;
RT "The pro-alpha3 (V) collagen chain. Complete primary structure,
RT expression domains in adult and developing tissues, and comparison to
RT the structures and expression domains of the other types V and XI
RT procollagen chains.";
RL J. Biol. Chem. 275:8749-8759(2000).
DB EMBL; AF177941; AAP53902.1; -.
DB InterPro; IPR000087; Collagen.
DB InterPro; IPR000885; Fib_collagen_C.
DB InterPro; IPR001791; Laminin_G.
DB InterPro; IPR003129; TSPN.
DB Pfam; PF01410; COL1F; 1.
DB Pfam; PF01391; Collagen; 17.
DB Pfam; PF02210; TSPN; 1.
DB ProDom; PD000007; Collagen; 3.
DB ProDom; PD002078; Fib_collagen_C; 1.
DB SMART; SM00038; COL1F; 1.
DB SMART; SM00282; LamG; 1.
DB SMART; SM00210; TSPN; 1.
SQ SEQUENCE 1745 AA; 172051 MW; 5E8FF97135397AC1 CRC64;

Query Match 7.8%; Score 145; DB 4; Length 1745;
Best Local Similarity 24.0%; Pred. No. 0.013;
Matches 88; Conservative 24; Mismatches 103; Indels 152; Gaps 20;

```

```

OY 174 GAVPRKPVSEMERDRGAHSLPEKENTLPDDPTSNATSRGABGLPDP-----PSNPA 227
DB 960 EUGPSPGKKEGPGALRGF-----PGPFGGPGDPGPTGL-KGDKPGPGVANGSPGERG 1013
OY 228 VAGAAGLALLLLGVAGAGAMCWRRRRAKPSSESHPG-PSFGGSLG-LGGGGGNGP 285
DB 1014 PLGPAGGIC-LPGSGSGSPVGPAGKKSRRGERPPGPTGKGDPGLGFLGPPGAAGP 1071
OY 286 --REAPGELGIALRG-----GAADPPCPHYEKSGDGHVYIVODGPP-----330
DB 1072 SGEEDKGVGAPGHKSGKXGDKDAGPPGCP---GIRGPAHPP-----GPPADGAQGR 1122
OY 331 GSPPNY 337
DB 1123 RGPGLF 1129

RESULT 30
O25466 PRELIMINARY: PRT, 309 AA.
AC O25466:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE COL-1.
GN COL-1.
OS Meloidogyme incognita (southern root-knot nematode).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchoidea; Heteroderidae; Meloidogyminae; Meloidogyne.
OX NCBI_TaxID=6306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97164124; PubMed=9010847;
RA Ray C., Wang T.Y., Husey R.S.;
RT "Identification and characterization of the Meloidogyme incognita coll
RT cuticle collagen gene.";
RL Mol. Biochem. Parasitol. 83:121-124(1996).
DR EMBL; U40766; AAC47437.1; -.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01391; Collagen_2.
DR Pfam; PF01484; Col_cuticle_N; 1.
SQ SEQUENCE 309 AA; 29472 MW; 0C6C9D5BA76C8916 CRC64;

Query Match 7.8%; Score 144.5; DB 5; Length 309;
Best Local Similarity 28.4%; Pred. No. 0.002;
Matches 46; Conservative 14; Mismatches 67; Indels 35; Gaps 6;

OY 171 PRGAVPRKPVSEMERDRGAHSLPEKENTLPDDPTSNATSRGABGLPDPSPNPAVAG 230
DB 156 PCPGCPGPGPPGPPGCGPGCGPGCGQASFGBP-----GPAGPPGSPGAPG 205
OY 221 AAGGLALLLLGVAGAGAMCWRRRRAKPSSESHPG---PSFGGSLG-LGGGGGNGP 285
DB 206 QGCG-----PGQPSGA-OSQAGAPAPPPGPDADAPGAPGAPGAGGPGGPGP 256
OY 286 R-----EAPGELGIALRGGAADPPCFPHYKVS 316
DB 257 KGPSPGPGNDGAPGPGTGTGAGAGKGIKPYCALDG 298

RESULT 31
O960B3 PRELIMINARY: PRT, 1347 AA.
AC O960B3:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-UTN-2002 (TREMBLrel. 21, Last annotation update)
DE Type V preprocollagen alpha 2 chain (Fragment).
GN COL5A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21451029; PubMed=11566270;
RA Valkila M., Melkonian M., Kvist L., Kuivaniemi H., Tromp G.,
RA Ala-Kokko U.;
RT "Genomic differentiation of the human COL3A1 and COL5A2 genes: COL5A2 has
RT evolved differently than the other minor fibrillar collagen genes.";
RL Matrix Biol. 20:357-366(2001).
DR EMBL; AY016295; AAL13166.1; -.
DR EMBL; AY016289; AAL13166.1; JOINED.
DR EMBL; AY016290; AAL13166.1; JOINED.
DR EMBL; AY016291; AAL13166.1; JOINED.
DR EMBL; AY016292; AAL13166.1; JOINED.
DR EMBL; AY016293; AAL13166.1; JOINED.
DR EMBL; AY016294; AAL13166.1; JOINED.
DR InterPro; IPR000085; Fibr_collagen_C.
DR InterPro; IPR000885; Fibr_collagen_C.
DR Pfam; PF01410; COLFi_1.
DR Pfam; PF01391; Collagen_18.
DR ProDom; PD000007; Collagen; 6.
DR ProDom; PD002078; Fibr_collagen_C; 1.
KW Collagen.
FT NON TER.
SQ SEQUENCE 1347 AA; 128822 MW; F25F20E6B23A17C CRC64;

Query Match 7.8%; Score 144.5; DB 4; Length 1347;
Best Local Similarity 33.0%; Pred. No. 0.011; 77; Indels 35; Gaps 11;
Matches 61; Conservative 12; Mismatches 77; Indels 35; Gaps 11;

OY 169 QSPRG--GAV-PRKPVSEMERDRGAHSLPEKENTLP-----DPTSNATSRGAE 217
DB 351 RGRGDPGTVGPPGPGVGEAGPKNRGF-----PGSDGLPGRGAGGERPVSSSGKXGQ 405
OY 218 GRLPPSPNPAVAGAGLALLLLGVAGAGAMCWRRRRAKPSSESHPG-PSFG-RG--G 273
DB 406 GDGRGPGEPGLPGAR-----LTGNPGVQPGPKLPLGAPGEDGPPGPGSIGRGPG 460
OY 274 SLGLGGGGGKGPPEAPGELGIA---LRGGGAADPPCFPHYKVSVDGHVYIVODG 329
DB 461 SMGLPGRKSSGDPGRKPGAGNAGVPGQAGAPKQGVSP-----SGVPFGLAGERG- 514
OY 330 PGSP 334
DB 515 EQGPP 519

RESULT 32
O90W37 PRELIMINARY: PRT, 1420 AA.
AC O90W37:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-UTN-2002 (TREMBLrel. 21, Last annotation update)
DE Alpha 1 type IIA collagen precursor.
GN COL2A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=STERNAL;
RA Caixia X., Yongzhi X., Siqi G., Yiyang S.;
RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY046949; AAK96621.1; -.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fibr_collagen_C.
DR Pfam; PF01410; COLFi_1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Collagen; 6.
DR ProDom; PD002078; Fibr_collagen_C; 1.

```

KM Collagen: signal.  
 FT SIGNAL 1 25 POTENTIAL.  
 SQ SEQUENCE 1420 AA; 13499 MW; 88D9AAB17F214FF5 CRC64;  
 Query Match  
 Best Local Similarity 28.1%; Score 144.5; DB 13; Length 1420;  
 Matches 65; Conservative 11; Mismatches 90; Indels 65; Gaps 10;

QY 147 GLESLQG--GVCILTRGMKVLRLVQSGFRGAVVRK-----VSEMER----- 188  
 DB 644 GAQGLQGRRLPGTPTG-----DGPKGATGPAAGNAGQPPGLQGMGEKRAAGIAGL 696  
 QY 189 --TRGAHSLPEPKENLPDPTSNATSRGAEPLPPSPMAVAGAGLALLLVAGAG 246  
 DB 697 KGRGVDVG--EKPEAGAG---KDGAKILTPITGPPGAPGVEKESG--PPPSGAA 748  
 QY 247 GANCMRRRAKPSSESRHPG--PGSFRGGSLGLGGGGMGPREAPBELGIALRG----- 300  
 DB 749 GA-----RGAPGERGPAFPAFPAFGADGQPGAKGEQEPQKGDAGAPQGPS 802  
 QY 301 -----GAADPPCEHYEKVSGDYGHPIYIVQDPPSP 333  
 DB 803 GADPGQPGTGTGPKAGAGQGPATGFPAGAGVGPQPGNPGPP 853

RESULT 33  
 Q9F342 PRELIMINARY; PRT; 775 AA.  
 AC Q9F342;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)  
 DE Hypothetical protein SC05128.  
 GN SC05128 OR SC9812.13.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.  
 NX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Lake L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S., Taylor K.,  
 RA Seeger T., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL: AL391751; CAC05758.1; -.  
 KM Hypothetical protein.  
 SQ SEQUENCE 775 AA; 75820 MW; E5032698A20A0711 CRC64;

Query Match  
 Best Local Similarity 7.8%; Score 143.5; DB 16; Length 775;  
 Matches 94; Conservative 23; Mismatches 130; Indels 153; Gaps 19;

QY 3 PPHSGP-----GVRVGLALLLGLVLSGLSLPEPYWNSANKRFOABEGYVL 50  
 DB 133 FRSGSGPKTQGGSGTNGAGLPAGSAPTAGTPAAGARGAGPAGAGAG----- 181  
 QY 51 YPGIGRLDLCPARPPGHSSPNVEFYKLYLVGAQO-----GRCFA---PPAPNLL 101  
 DB 182 -----ARGGSGSRPGSVGSMANPGARSSAGGARGATGFPVP--- 224  
 QY 102 LTCRDLRLRFTTKFQEGYSPNLMGHEFRSHHDYIIAIVSDGTREGLESLQGVCLTRGM 161  
 DB 225 -----GH-----GGGTGSFDTVEALLAGPPLGNGN 249

QY 162 KYLLRVGQSP-----GAIVPRKPVSEMPMERDRGAHSLBPKENLPGDPTSNATSRGA 217  
 DB 250 GNGNGNGSGPFPPTDAGGEBPR--DDLFPAGEBRAAOAPGQGN--GLNGQGSYGTGPQ 305  
 QY 218 GPLPPSPWPAV-----AGAAGLALLLVGAAGAGANCMWRRAKPSSESRHP----- 264  
 DB 306 GPAPPTSPAPAGDSRLRTPPAGDGLG-----LGTAGGPAA-----POTQSEPAFNG 351  
 QY 265 --PGPSF--GRGSLGLGGGGMGPREAPBELGIALRGGAAD-----PPPCPH 310  
 DB 352 LRGPSTLGGPSPGPGPAGPAGSGPDAARLGP-----GGGISDPTAILTPQRPVPPGAN 405  
 QY 311 YEKVSGD---YGHPIYIVQ-----DGP--QSPENI 336  
 DB 406 PDNISGNTVTSIGPIVVPGERTAPPAGSGDGPHTPPKL 445

RESULT 34  
 Q26640 PRELIMINARY; PRT; 1747 AA.  
 AC Q26640;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)  
 DE Alpha2 (IV)-like collagen.  
 GN COLP4ALPHA.  
 OS Strongylocentrotus purpuratus (Purple sea urchin).  
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;  
 OC Strongylocentrotus.  
 NX NCBI\_TaxID=7668;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94230414; PubMed=8175744;  
 RA Expósito J.Y., Suzuki H., Geourjon C., Garrone R., Solursh M.,  
 RA Ramirez F.;  
 RT "Identification of a cell lineage-specific gene coding for a sea  
 urchin, alpha.2(IV)-like collagen chain.";  
 RL J. Biol. Chem. 269:13167-13171(1994).  
 DR EMBL: X76730; CAA54146.1; -.  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR001442; ProcollagenC4.  
 DR Pfam: PF01413; C4; 2.  
 DR Pfam: PF01391; Collagen; 21.  
 DR ProDom: PD000007; Collagen; 6.  
 DR ProDom: PD003923; ProcollagenC4; 1.  
 DR SMART: SM00111; C4; 2.  
 KM Collagen.  
 SQ SEQUENCE 1747 AA; 173312 MW; EE722E878394B986 CRC64;

Query Match  
 Best Local Similarity 7.7%; Score 143; DB 5; Length 1747;  
 Matches 84; Conservative 24; Mismatches 159; Indels 82; Gaps 14;

QY 7 GPGYRVGALLLGLVLSGLSLPEPYWNSANKRFOABEGYVLPPIGRLDLCPRAR 66  
 DB 266 GPRGMDG---MKATGEVGLDLSGYDYGEKGLPGYSGERGFPNPPIGLGNNGEKGR 321  
 QY 67 PPGHSSPNVEFYKLYLVGAQGRCEAPAPNLLITCDRPDLRFTIKFQEYSPNLMG 126  
 DB 322 -DGQGRGNYGK--GPSGDYGMDDGPAPDL----- 352  
 QY 127 HEFRSHHDYIIAIVSDGTREGLESLQGVCLTRGMKYLRLVQO---SPRGAVPRKPV 182  
 DB 353 -----EELVGVGPDPDEGPPNPGRPSGALGLFGQGGYGPMPGPPRPPTG 405  
 QY 183 EMPMERDRGAHSLBPKENLPDPTSNATSRGAEPLPPSPMAVAGAA 232  
 DB 406 SGGRGRDSEKESRPGTPTGSETGRANDGNFGEKGMWESRPGPMDGSTRGQ 465  
 QY 233 GGLALLLVGAAGAGANCMWRRAKPSSESRHP-----PGSFRGSLGLGGGGMGP 285  
 DB 466 G-----FWQKGRGRP-----PGRAGPAGPARSGNQSFQGFPGDTGNTGLKGMGIQALG 516

QY 286 REAFPELIGALRGGAADPPCPHYEK-VSGDGHPPYIVQDGPQSP 333  
 DB 517 RDGRGSKG---ELGIGC-PPCPGKGYVPGDRGVP-----GDPGSP 554

RESULT 35

ID Q96D07 PRELIMINARY; PRT; 744 AA.  
 AC Q96D07;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 OS Hypothetical 73.4 kDa protein.  
 DE Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LUNG;  
 RA Strusberg R.;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC013581; AAH13581.1; -  
 DR InterPro; IPR001073; C1Q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1Q; 1.  
 DR Pfam; PF01391; Collagen; 7.  
 DR ProDom; PD000007; Collagen; 2.  
 DR PROSITE; PS0113; C1Q; UNKNOWN\_1.  
 KM Collagen; Hypothetical protein.  
 SQ SEQUENCE 744 AA; 73364 MM; 2BC1B0955DE2C9A3 CRC64;

Query Match 7.7%; Score 142; DB 4; Length 744;  
 Best Local Similarity 23.3%; Pred. No. 0.0084;  
 Matches 72; Conservative 16; Mismatches 99; Indels 122; Gaps 13;

QY 67 PGPSPSPHYEFYKLYVGAGAGRCAPAPNLLTCDRPLDLRTFKGEVSPNLMG 126  
 DB 198 PGPSPSPHYEFYKLYVGAGAGRCAPAPNLLTCDRPLDLRTFKGEVSPNLMG 222  
 QY 127 HFRSHHDYIATSDGTREGLSELOGGVCLTRGMKVLRLVQSGPRGAVPRKPYSEMPM 186  
 DB 223 -----GDRGPKGLPGPQ-----LRGPKGKGCWMPGAPGVKGP- 256  
 QY 187 EMDRGAASLEPKENLPDPTSNAT-SRGAEPFPPSPRAVAVAGAGLLLL-----G 241  
 DB 257 ---PGRHG-PPGPVGLPGVGKPGVTGPPGPGPLGKPGAPGEPGPGPIGVPGVQPPG 311  
 QY 242 VAGAGGAMCWRRAKPSRHPG-----PGSFGGRGS 274  
 DB 312 IIGIG-----KPGDDGIPGPGPPGPGKGEGLPLGPPGRLPGIKGPPGPPGPD 361  
 QY 275 LGLGG-GGGMGRREAPFGLGIALRGGAADPPCFHYEKVSGDYGHVYIVQDGP----- 329  
 DB 362 RGMGVPGFALGPR-GEKPGIGAPGIGG---PPGEPGLPGIPGMPGPGALGPGPKGEG 416  
 QY 330 ----PQSP 334  
 DB 417 GIVPGQSP 425

RESULT 36

ID Q91AU4 PRELIMINARY; PRT; 1835 AA.  
 AC Q91AU4;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 OS Alpha 1 (V) collagen.  
 DE Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RED WHITE LEGHORN;  
 RX MEDLINE=20068042; PubMed=10601735;  
 RA Gordon M.K., Marchant J.K., Foley J.W., Igoe F., Gibney E.P.,  
 RA Nah H.D., Barembaum M., Myers J.C., Rodriguez E., Dublet B.,  
 RA van der Rest M., Linsmeayer T.F., Upholt W.B., Birk D.E.;  
 RT "Complete primary structure of the chicken alpha1(V) collagen chain."  
 RL Matrix Biol. 18:481-486(1999).  
 DR EMBL; AF137273; AAF28099.1; -  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR000885; F1D\_collagen\_C.  
 DR InterPro; IPR01791; Lamlnin G.  
 DR InterPro; IPR01230; Prenyl\_site.  
 DR InterPro; IPR003129; TSPN.  
 DR Pfam; PF01410; COLFI; 1.  
 DR Pfam; PF01391; Collagen; 19.  
 DR Pfam; PF02210; TSPN; 1.  
 DR ProDom; PD000007; Collagen; 2.  
 DR ProDom; PD002078; F1D\_collagen\_C; 1.  
 DR SMART; SM0038; COLF; 1.  
 DR SMART; SM00282; Lamg; 1.  
 DR SMART; SM00210; TSPN; 1.  
 DR PROSITE; PS00294; PRENYLATION; UNKNOWN\_1.  
 KM Collagen.  
 SQ SEQUENCE 1935 AA; 184234 MM; D05B9D71022D44B2 CRC64;

Query Match 7.7%; Score 142; DB 13; Length 1835;  
 Best Local Similarity 25.7%; Pred. No. 0.024;  
 Matches 78; Conservative 15; Mismatches 82; Indels 128; Gaps 18;

QY 68 PGPSPSPHYEFYKLYVGAGAG-----RCCEAPAPNLLTCDRPLDLRTFKF 117  
 DB 566 PGPSPSPHYEFYKLYVGAGAG-----SGDIKGEVEMGPQGPRIQGPSP-----AGP 602  
 QY 118 QEYSPMLGHEPSSHHDYIATSDGT-----AGSDGARMPGQTPGKGRGDLGALPG 636  
 DB 603 -----GRGR-----AGSDGARMPGQTPGKGRGDLGALPG 636  
 QY 164 LRLVQSGPRGAVPRKPYSEMPERDRGAASLEPKENLPDPTSNATSRGAEGLPFP 223  
 DB 637 ---EKGNRGPBPDPGPPGPPGDEGRG--DDGEVPRGLPGSP-----GPRGLGPKGP 686  
 QY 224 SMPRAVAGAGLLLLLVAGAGAMCWRRAKPSRHPG---PGSFGGSLGI--GGGG 281  
 DB 687 GPPGPPVAG---WDGTPGKGNV-----GPGEPGPPGQGPAGAGLPDPG 732  
 QY 282 GMPRAEPFGLGIALRG-----GADPPCFHYEKVSGDYGHVYIVQDGP-----Q 331  
 DB 733 PIGP---PGEKPRLPKPLGPGMPGADGP-----GHP---GKEGPPGKSGQ 773  
 QY 332 SP 334  
 DB 774 GPP 776

RESULT 37

ID Q25582 PRELIMINARY; PRT; 284 AA.  
 AC Q25582;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Cuticular collagen.  
 GN COLOST-1.  
 OS Teladorsagia circumcincta.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdida; Strongylida;  
 OC Trichostrongyloidea; Haemonchidae; Oestertaglinae; Teladorsagia.  
 OX NCBI\_TaxID=45464;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=97039674; PubMed=8885225;  
 RA Johnstone I.L., Shafi Y., Mjaaid A., Barry J.D.;  
 RT "Cuticular collagen genes from the parasitic nematode *Ostertagia circumcincta*.";  
 RL Mol. Biochem. Parasitol. 80:103-112(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Johnstone I.L.;  
 RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20017523; PubMed=1051361;  
 RA Britton C., Redmond D.L., Knox D.P., McKerrrow J.H., Barry J.D.;  
 RT "Identification of promoter elements of parasite nematode genes in transgenic *Caenorhabditis elegans*.";  
 RL Mol. Biochem. Parasitol. 103:171-181(1999).  
 DR EMBL; X96731; CA65506.1; -;  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF01391; Collagen; 3.  
 DR Pfam; PF01484; Col\_cuticle\_N; 1.  
 KM Collagen.  
 SQ SEQUENCE 284 AA; 27534 MW; 60775B7D2CC40C77 CRC64;  
 Query Match 7.6%; Score 141; DB 5; Length 284;  
 Best local similarity 27.2%; Pred. No. 0.0033;  
 Matches 61; Conservative 15; Mismatches 86; Indels 62; Gaps 12;  
 QY 121 SPNMGHFRSHHDYIATSDGTREGLESLOGGVCLTRGMKVLRYVQSPRGAVPRKP 180  
 DB 62 SGNMREVTRTQ---VLAAP--TRKARQS-GGGCC-----GGGVSPAGPPGP--- 102  
 QY 181 VSEMPERDRGAHSLPEKKNLPDPTSNATSRGAGPLP-----PPSMPAVAGAA 232  
 DB 103 -----PGDQGPDSGDRGQPGRNPGDPATPAPDVAPCFNCPPGPPAGAP 151  
 QY 233 GGLALLLGVAGAGAMCRRRAKPSRSRHP-GPGSFRGSGSLGLGGGGMGPRAEPG 291  
 DB 152 GG-----RGQGPSGSS-----DQPGNAGNPGGPGPIGPFGPPGNAQPG--NAGSPG 197  
 QY 292 ELGILRGGAADPPFCPHYEKVSGDYGHVYIVODGPPGSPN 335  
 DB 198 AFGTLTGPGSGGPPGP-----GPGGP--GPDGQPGGPGN 232

RESULT 38  
 Q25581  
 ID Q25581 PRELIMINARY; PRT; 284 AA.  
 AC Q25581;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Cuticular collagen.  
 GN COL087-2.  
 OS Teladorsagia circumcincta.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongyloidea;  
 OC Trichostrongylidae; Haemonchidae; Ostertagiinae; Teladorsagia.  
 OC NCBI\_TaxID=54564;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97039674; PubMed=8885225;  
 RA Johnstone I.L., Shafi Y., Mjaaid A., Barry J.D.;  
 RT "Cuticular collagen genes from the parasitic nematode *Ostertagia circumcincta*.";  
 RL Mol. Biochem. Parasitol. 80:103-112(1996).  
 DR EMBL; X96731; CA65507.1; -;  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF01391; Collagen; 3.  
 DR Pfam; PF01484; Col\_cuticle\_N; 1.  
 KM Collagen.  
 SQ SEQUENCE 284 AA; 27560 MW; 7B7EB6D2DD41D76 CRC64;

Query Match 7.6%; Score 141; DB 5; Length 284;  
 Best local similarity 27.2%; Pred. No. 0.0033;  
 Matches 61; Conservative 15; Mismatches 86; Indels 62; Gaps 12;  
 QY 121 SPNMGHFRSHHDYIATSDGTREGLESLOGGVCLTRGMKVLRYVQSPRGAVPRKP 180  
 DB 62 SGNMREVTRTQ---VLAAP--TRKARQS-GGGCC-----GGGVSPAGPPGP--- 102  
 QY 181 VSEMPERDRGAHSLPEKKNLPDPTSNATSRGAGPLP-----PPSMPAVAGAA 232  
 DB 103 -----PGDQGPDSGDRGQPGRNPGDPATPAPDVAPCFNCPPGPPAGAP 151  
 QY 233 GGLALLLGVAGAGAMCRRRAKPSRSRHP-GPGSFRGSGSLGLGGGGMGPRAEPG 291  
 DB 152 GG-----RGQGPSGSS-----DQPGNAGNPGGPGPIGPFGPPGNAQPG--NAGSPG 197  
 QY 292 ELGILRGGAADPPFCPHYEKVSGDYGHVYIVODGPPGSPN 335  
 DB 198 AFGTLTGPGSGGPPGP-----GPGGP--GPDGQPGGPGN 232

RESULT 39  
 Q17038  
 ID Q17038 PRELIMINARY; PRT; 319 AA.  
 AC Q17038;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical 31.3 kDa protein.  
 GN T1587.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OC NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=8851916;  
 RA None;  
 RT "Genome sequence of the nematode *C. elegans*: a platform for investigating biology. The *C. elegans* Sequencing Consortium.";  
 RT Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Pauley A., Gattung S.;  
 RT "The sequence of *C. elegans* cosmid T1587.";  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Waterston R.;  
 RT "Direct Submission.";  
 RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF022985; AAB69959.1; -;  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR002486; Col\_cuticle\_N.  
 DR Pfam; PF01391; Collagen; 3.  
 DR Pfam; PF01484; Col\_cuticle\_N; 1.  
 KM Hypothetical protein.  
 SQ SEQUENCE 319 AA; 31292 MW; 5242A8C668A488 CRC64;  
 Query Match 7.6%; Score 141; DB 5; Length 319;  
 Best local similarity 25.6%; Pred. No. 0.0036;  
 Matches 90; Conservative 27; Mismatches 111; Indels 124; Gaps 23;  
 QY 13 VGALLLGVGLVSGLSI-----EPYVNSANKRFGABGG---VYLYPGIGRLDLC 62  
 DB 1 MSASTLVVAASAGIAIVCVFTVGMIFNDINSFYDEKIGELAEFKGQEIA--WQAMI 58  
 QY 63 PRAPPGHSSPNVEFYKLYVG-----GAQRRCBA--PPANNLLTCRPPDL 111  
 DB 59 PTPRPSGSS-----FLGNNKQAEKNCGEQSGRGCPAGPPGP-----GQPG--- 101

```

QY 112 RTTFFQYSPNLMGHEFRSHHDYITATSDGTREGLSLOGVCLTRGMKVLIRVQSP 171
D 102 ---ARGAGLPLFIAGQ-----PGSGARINPAGTGPFCIT-----CP 135
QY 172 RGGAIVRKPVSMPMERDRGAHSLPEPKENLPDPTSNATSRGAEGLPPSPMPAVAGA 231
D 136 AGAPRPAAP-----PGA-----PPKGNNGQPAPAPAS-GGRP-PGRPP-AGD 177
QY 232 AGGALLLIGVAGAGAMCWRRRRAKPSRHPG-PGSFGRG-SLGLGGGGG---MG 284
D 178 AG-----SPGQPGPGSPGNCRGGRSRRTPGASGRPGPG 214
QY 285 PREAPRGLALRGGAADP-PFCPHYKXSGDYGHVYIVODGPPQSPPN 335
D 215 PAGA-PGQPG---RSGGAGTPOPGP-----PGSPGPGHSGNDVPRTPN 257

```

## RESULT 40

```

ID 093485 PRELIMINARY; PRT; 809 AA.
AC 093485.
DT 01-NOV-1998 (TRENBLrel. 08, Created)
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Alpha 1 type I collagen (Fragment).
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Proactinopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FIBROBLAST;
RA Saito M., Kimura S., Hirono I., Aoki T., Ishida M., Urano N.,
RA Kimura S.,
RT Partial characterization of cDNA clones encoding the three distinct
RT pro alpha chains of type I collagen from rainbow trout.";
RL Fisheries Sci. 64:780-786(1998).
DR EMBL; AB008373; BAA3380.1; -.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 9.
DR ProDom; PD000007; Collagen; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
KW Collagen.
KW NON TER.
FT SEQUENCE 809 AA; 78164 MW; 68C056A7640FCAB1 CRC64;

```

Query Match 7.6%; Score 141; DB 13; Length 809;  
 Best Local Similarity 27.4%; Pred. No. 0.011;  
 Matches 61; Conservative 15; Mismatches 93; Indels 54; Gaps 9;

```

QY 139 ATSDGTREGLSLOG-----GVCLTRGMKVLIRVQSPRGAVRKPVSMPMERDRGAA 193
D 98 ADGAGCKGVRMTGPIGNPAGSPGDK---GETAPAVAGPSGARAGPGRGESGA 152
QY 194 HSLPEKKNLPDPTSNAT-----SRGAEGLPPSPMPAVAGAGLALLL 239
D 153 ---PGPAGFAPPGDGDGPAKGEAGDNGAKGDGAGAPGAGTGAPEQPAQN----- 203
QY 240 LGVAAGAGAMCWRRRRAKPSRHPG-----PSFGRGSLGLGGGGGMPREAPGE 292
D 204 TGAKARGA-----AGPFGATGFGAAGRFPGPSNNPPTGPGGKGKGNRGE 257
QY 293 LGIALRGG--GAADPPFCPHYKXSGDYGHVYIVODGPPQSP 333
D 258 TGPAGRPGLGAAGP-----GPXGKGPQGDGPNPSPGTP 294

```

## RESULT 41

```

Q910C0
ID Q910C0 PRELIMINARY; PRT; 1449 AA.
AC Q910C0;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Collagen a1(I).
GN COL1A1.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Proactinopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=2157802; PubMed=11358497;
RA Saito M., Takenouchi Y., Kimura S.,
RT "Complete primary structure of rainbow trout type I collagen
RT consisting of a1(I)a2(I)a3(I) heterotrimer.";
RL Eur. J. Biochem. 268:2817-2827(2001).
DR EMBL; AB052835; BABS5661.1; -.
DR InterPro; IPR000887; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; Collagen; 2.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR PROSITE; PS01208; VWF_C; UNKNOWN_1.
DR PROSITE 1449 AA; 137117 MW; 62E8F8A7BFD652B8 CRC64;

```

Query Match 7.6%; Score 141; DB 13; Length 1449;  
 Best Local Similarity 27.4%; Pred. No. 0.022;  
 Matches 61; Conservative 15; Mismatches 93; Indels 54; Gaps 9;

```

QY 139 ATSDGTREGLSLOG-----GVCLTRGMKVLIRVQSPRGAVRKPVSMPMERDRGAA 193
D 738 ADGAGCKGVRMTGPIGNPAGSPGDK---GETGAPAVGSGARAGPGRGESGA 792
QY 194 HSLPEKKNLPDPTSNAT-----SRGAEGLPPSPMPAVAGAGLALLL 239
D 793 ---PGPAGFAPPGDGDGPAKGEAGDNGAKGDGAGAPGAGTGAPEQPAQN----- 843
QY 240 LGVAAGAGAMCWRRRRAKPSRHPG-----PSFGRGSLGLGGGGGMPREAPGE 292
D 844 TGAKARGA-----AGPFGATGFGAAGRFPGPSNNPPTGPGGKGKGNRGE 897
QY 293 LGIALRGG--GAADPPFCPHYKXSGDYGHVYIVODGPPQSP 333
D 898 TGPAGRPGLGAAGP-----GPXGKGPQGDGPNPSPGTP 934

```

## RESULT 42

```

ID Q90YC5 PRELIMINARY; PRT; 219 AA.
AC Q90YC5;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Ephrin-A3.
GN EPHRIN-A3.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21412237; PubMed=11520665;
RA Hirate Y., Mieda M., Harada T., Yamasu K., Okamoto H.;
RT "Identification of ephrin-A3 and novel genes specific to the midbrain-
RT MB in embryonic zebrafish by ordered differential display.";

```



```

RESULT 45
ID Q97405 PRELIMINARY; PRT; 1378 AA.
AC Q97405;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Collagen pro alpha-chain precursor.
GN HCOL 1 ALPHA.
OS Haliotis discus (Abalone).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Archaeogastropoda;
CC Haliotidae; Haliotis.
CX NCBI_TaxID=36094;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE;
RX MEDLINE=99234051; PubMed=10215888;
RA Yoneda C., Hirayama Y., Nakaya M., Matubara Y., Irie S., Hatae K.,
  Matabe S.;
RT "The occurrence of two types of collagen proalpha-chain in the abalone
  Haliotis discus muscle.";
RL Eur. J. Biochem. 261:714-721(1999).
DR EMBL; AB017600; BAA75668.1; -
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 17.
DR ProDom; PD000007; Collagen; 1.
DR ProDom; PD002078; Fib collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR Signal.
KM SIGNAL.
SQ SEQUENCE 1378 AA; 132583 MW; 6EEB34A65B52E634 CRC64;
  1 18 POTENTIAL.
Query Match 7.6%; Score 140; DB 5; Length 1378;
Best Local Similarity 25.4%; Pred. No. 0.024;
Matches 96; Conservative 21; Mismatches 125; Indels 136; Gaps 20;
QY 2 GPHSGPGGVV-VGALLLVGLVGLSLPEPYNN--SANKRFQAEQGVLYLPQIGDR 57
DB 646 GPP--GPPAGAGDAAGGLGMPG---ERPGIRNGPQGNRLTGERG---QDGE- 692
QY 58 LDLCPARP-----PGHSSPNYEFYLYVGAQGRCEAPAPNLLTCDR 106
DB 693 -----PGRPEBAGAPGSPQPGPSG-----LVSAKDRGRAGPA----- 726
QY 107 PDLRLFTKFEYSPLNMGHEFRSHHDYIIATSDGTR--EGLESIQGVCLTRGMKVL 164
DB 727 -----GEPGP-----PGSSGGRGAPAGQPGPTGSLGPTGEM 759
QY 165 LRVGSPGAGAVPRKPVSEMERDRGAASLE-----PGKENLPDGP--T 208
DB 760 GQTSGDGDGAKGTGAGCYFGEAGPICAPNENGRBKRSRSGGIPGNSGTPGDPGRA 819
QY 209 SNATSRGAEGPLPPSPMPAVAGAAAGLALLLVGAAGAGAMCWRRAKPSSESHPG-- 265
DB 820 GPPGSPGAGQGP-----PGSPGAT-----GLSGDGE--RGETGPPGRSGEPGAPG 862
QY 266 -----PGSGRGSGSLGIGGGG--MGPR-----EAEPGELGILRGGADPFCFHYE 312
DB 863 MPGLDGAQGERGSPGLNPGSPGPPGVGFQGERGANGFPGSGEA---GAAGPSSAGGP 918
QY 313 KVSQDYGHVYIVDGP 330
DB 919 GLRGDNGNDGAPGQAGP 936
RESULT 46
ID Q9Y1B4 PRELIMINARY; PRT; 1450 AA.

```

```

AC Q9Y1B4;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Alpha 1 type I collagen.
OS Cynops pyrrhogaster (Japanese common newt).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.
CX NCBI_TaxID=8330;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=REGENERATE FORELIMBS;
RX MEDLINE=99407244; PubMed=10474166;
RA Asahina K., Obara M., Yoshizato K.;
RT "Expression of genes of type I and type II collagen in the formation
  and development of the blastema of regenerating newt limb.";
RL Dev. Dyn. 216:59-71(1999).
DR EMBL; AB015438; BAA36973.1; -
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Collagen; 1.
DR ProDom; PD002078; Fib collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C; UNKNOWN_1.
KM Collagen.
SQ SEQUENCE 1450 AA; 137564 MW; ABF8A74841B97B7C CRC64;
  1 18 POTENTIAL.
Query Match 7.5%; Score 139.5; DB 13; Length 1450;
Best Local Similarity 25.7%; Pred. No. 0.028;
Matches 103; Conservative 24; Mismatches 137; Indels 137; Gaps 24;
QY 2 GPHSGPGGVV-VGALLLVGLVGLSLPEPYNN--SANKRFQAEQGVLYLPQIGDR 59
DB 552 GPP--GPPAGAGQGS---GVMGFPQPKGAGP-----GKSGER-- 584
QY 60 LLCPRAPPPGPHSSPNYEFYLYVGAQGRCEAPAPNLLTCDRDLRLFTKFE 119
DB 585 ---GVAGPPGATGAKCKGQ-----EAGAQG-----PPPS-CPGSEKREQGPASSPGQG 630
QY 120 Y--SPNLWGEFRSHHDYIIATSDGTREGLESIQG--GVCLTRGMKVLVVGSPRGA 175
DB 631 LPGSPGAPGAGAKPGEQ--APGDAGGPPGSPGRGERGFGERG-----GQGFPAQ 680
QY 176 VPR---KPVSE-----MPMER-----DRGAHSLPBG 199
DB 681 GPPGSPGSPGNDGAKGEMAGAGAPGGRPPGLQGMPPGERGSAKMPGAKGDKGA-----G 735
QY 200 KENLPDPTSNATSRGAEGLPPSPMPAVAG--AAGLALLLVGAAGAMCWRRAK 257
DB 736 TGADGAPGKDG-ARGTLGPICPPGSPGAPDKGEGPS---GAGPTGA---RGS 784
QY 258 PESRHPG-----PGSFRGSGSLG---LGGGGMGRREAPGELGIAL 297
DB 785 PGRGEPGAPGAGICGPPGADGQPGAGKESGDAFPKDAAPGAPGPTGA-PGPAAGV 843
QY 298 RGG-----GADPPFCPHYEKVSGDYGHVYIVDGPDPSP 333
DB 844 APGPKGTGGAAGFPAGATGPPGAAGRLGPPGSPGNAAGPFGP 884
RESULT 47
ID Q9UMG6 PRELIMINARY; PRT; 1690 AA.
AC Q9UMG6;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Collagen type IV a6 chain.
GN COL4A6.

```



OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OK NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE OF 1-1644 FROM N.A.  
 RX MEDLINE=9629642; PubMed=8661006;  
 RA Zhang X., Zhou J., Reeders S.T., Tryggvason K.;  
 RT "Structure of the human type IV collagen COL4A6 gene, which is mutated  
 in Alport syndrome-associated leiomyomatosis.";  
 RL Genomics 33:473-479(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Zhang X.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL U47004; AAB19039.1; -. JOINED.  
 DR EMBL U46960; AAB19039.1; JOINED.  
 DR EMBL U46961; AAB19039.1; JOINED.  
 DR EMBL U46962; AAB19039.1; JOINED.  
 DR EMBL U46963; AAB19039.1; JOINED.  
 DR EMBL U46964; AAB19039.1; JOINED.  
 DR EMBL U46965; AAB19039.1; JOINED.  
 DR EMBL U46966; AAB19039.1; JOINED.  
 DR EMBL U46967; AAB19039.1; JOINED.  
 DR EMBL U46968; AAB19039.1; JOINED.  
 DR EMBL U46969; AAB19039.1; JOINED.  
 DR EMBL U46970; AAB19039.1; JOINED.  
 DR EMBL U46971; AAB19039.1; JOINED.  
 DR EMBL U46972; AAB19039.1; JOINED.  
 DR EMBL U46973; AAB19039.1; JOINED.  
 DR EMBL U46974; AAB19039.1; JOINED.  
 DR EMBL U46975; AAB19039.1; JOINED.  
 DR EMBL U46976; AAB19039.1; JOINED.  
 DR EMBL U46977; AAB19039.1; JOINED.  
 DR EMBL U46978; AAB19039.1; JOINED.  
 DR EMBL U46979; AAB19039.1; JOINED.  
 DR EMBL U46980; AAB19039.1; JOINED.  
 DR EMBL U46981; AAB19039.1; JOINED.  
 DR EMBL U46982; AAB19039.1; JOINED.  
 DR EMBL U46983; AAB19039.1; JOINED.  
 DR EMBL U46984; AAB19039.1; JOINED.  
 DR EMBL U46985; AAB19039.1; JOINED.  
 DR EMBL U46986; AAB19039.1; JOINED.  
 DR EMBL U46987; AAB19039.1; JOINED.  
 DR EMBL U46988; AAB19039.1; JOINED.  
 DR EMBL U46989; AAB19039.1; JOINED.  
 DR EMBL U46990; AAB19039.1; JOINED.  
 DR EMBL U46991; AAB19039.1; JOINED.  
 DR EMBL U46992; AAB19039.1; JOINED.  
 DR EMBL U46993; AAB19039.1; JOINED.  
 DR EMBL U46994; AAB19039.1; JOINED.  
 DR EMBL U46995; AAB19039.1; JOINED.  
 DR EMBL U46996; AAB19039.1; JOINED.  
 DR EMBL U46997; AAB19039.1; JOINED.  
 DR EMBL U46998; AAB19039.1; JOINED.  
 DR EMBL U46999; AAB19039.1; JOINED.  
 DR EMBL U47000; AAB19039.1; JOINED.  
 DR EMBL U47001; AAB19039.1; JOINED.  
 DR EMBL U47002; AAB19039.1; JOINED.  
 DR EMBL U47003; AAB19039.1; JOINED.  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR001442; ProcollagenC4.  
 DR Pfam: PF01413; C4; 2.  
 DR Pfam: PF01391; Collagen; 23.  
 DR ProDom: PD000007; Collagen; 4.  
 DR ProDom: PD003923; ProcollagenC4; 1.  
 DR SMART: SM00111; C4; 2.  
 KW Collagen.  
 SQ SEQUENCE 1690 AA; 163696 MW; 4698AE2CC2D3E859 CRC64;

QY 147 GLESLQ--GYCTRMKTVL-----RVGSPRGA-----VPRKPYSEM 184  
 Db 905 GFPGIPGLPGISGRILKIKGPGSTGKMGPSGRAGTPEKCDRGNPGVGI:PSRRPMSNL 964  
 QY 185 PMERDRGAASL-----EPGKENTLPDDPTSNATSRQAG-PLP----- 222  
 Db 965 WLKDKSGSQSGASNGRPFGRGDKGEGRPGPGLGAPPLPGLIKVSGKPGPFGMT 1024  
 QY 223 -----PMPVAAGA-----GL--ALLIGVAGAGAMCRRRRAXPSE 260  
 Db 1025 RGLPGLKSGSGITGFPMPMPESGQGRSGPLPGASGLPGLKGNQGVTV--EISGSPGP 1082  
 QY 261 SRHPGGSF--GSGSLGLGGCGMGPREAPRGLGIALRGGAADPPGCHYKVSQDVG 319  
 Db 1083 KGQPSGSGFKYGRDGLIGNIGPFGKKGDKVGS---GVLGLGAPGPGVAGNRG 1138  
 QY 320 HP 321  
 Db 1139 EP 1140  
 RESULT 48  
 ID QY4L4 PRELIMINARY; PRT; 1691 AA.  
 AC QY4L4;  
 DT 01-NOV-1999 (T-EMBLrel. 12, Created)  
 DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
 GN Collagen type IV a6 chain.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OK NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE OF 1-1645 FROM N.A.  
 RX MEDLINE=9629642; PubMed=8661006;  
 RA Zhang X., Zhou J., Reeders S.T., Tryggvason K.;  
 RT "Structure of the human type IV collagen COL4A6 gene, which is mutated  
 in Alport syndrome-associated leiomyomatosis.";  
 RL Genomics 33:473-479(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Zhang X.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL U47004; AAB19038.1; -. JOINED.  
 DR EMBL U46959; AAB19038.1; JOINED.  
 DR EMBL U46960; AAB19038.1; JOINED.  
 DR EMBL U46961; AAB19038.1; JOINED.  
 DR EMBL U46962; AAB19038.1; JOINED.  
 DR EMBL U46963; AAB19038.1; JOINED.  
 DR EMBL U46964; AAB19038.1; JOINED.  
 DR EMBL U46965; AAB19038.1; JOINED.  
 DR EMBL U46966; AAB19038.1; JOINED.  
 DR EMBL U46967; AAB19038.1; JOINED.  
 DR EMBL U46968; AAB19038.1; JOINED.  
 DR EMBL U46969; AAB19038.1; JOINED.  
 DR EMBL U46970; AAB19038.1; JOINED.  
 DR EMBL U46971; AAB19038.1; JOINED.  
 DR EMBL U46972; AAB19038.1; JOINED.  
 DR EMBL U46973; AAB19038.1; JOINED.  
 DR EMBL U46974; AAB19038.1; JOINED.  
 DR EMBL U46975; AAB19038.1; JOINED.  
 DR EMBL U46976; AAB19038.1; JOINED.  
 DR EMBL U46977; AAB19038.1; JOINED.  
 DR EMBL U46978; AAB19038.1; JOINED.  
 DR EMBL U46979; AAB19038.1; JOINED.  
 DR EMBL U46980; AAB19038.1; JOINED.  
 DR EMBL U46981; AAB19038.1; JOINED.  
 DR EMBL U46982; AAB19038.1; JOINED.  
 DR EMBL U46983; AAB19038.1; JOINED.  
 DR EMBL U46984; AAB19038.1; JOINED.  
 DR EMBL U46985; AAB19038.1; JOINED.

Query Match 7.5%; Score 139.5; DB 4; Length 1690;  
 Best Local Similarity 25.2%; Pred. No. 0.034;  
 Matches 61; Conservative 25; Mismatches 83; Indels 73; Gaps 11;

```

DR EMBL; U46986; AAB19038.1; JOINED.
DR EMBL; U46987; AAB19038.1; JOINED.
DR EMBL; U46988; AAB19038.1; JOINED.
DR EMBL; U46989; AAB19038.1; JOINED.
DR EMBL; U46990; AAB19038.1; JOINED.
DR EMBL; U46991; AAB19038.1; JOINED.
DR EMBL; U46992; AAB19038.1; JOINED.
DR EMBL; U46993; AAB19038.1; JOINED.
DR EMBL; U46994; AAB19038.1; JOINED.
DR EMBL; U46995; AAB19038.1; JOINED.
DR EMBL; U46996; AAB19038.1; JOINED.
DR EMBL; U46997; AAB19038.1; JOINED.
DR EMBL; U46998; AAB19038.1; JOINED.
DR EMBL; U46999; AAB19038.1; JOINED.
DR EMBL; U47000; AAB19038.1; JOINED.
DR EMBL; U47001; AAB19038.1; JOINED.
DR EMBL; U47002; AAB19038.1; JOINED.
DR EMBL; U47003; AAB19038.1; JOINED.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001442; ProcollagenC4.
DR Pfam; PF04113; C4; 2.
DR Pfam; PF01391; Collagen; 23.
DR ProDom; PD000007; Collagen; 4.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
SQ SEQUENCE 1691 AA; 163873 MW; 769AA53D1C7CA87 CRC64;

Query Match 7.5%; Score 139.5; DB 4; Length 1691;
Best Local Similarity 25.2%; Pred. No. 0.034;
Matches 61; Conservative 25; Mismatches 83; Indels 73; Gaps 11;

QY 147 GLESLQG--GYCLTSGMKVLL---RVGSGPRGA-----VPRKPYSEM 184
DB 906 GPPGIPGLPGISGTRGLKGLPGSTGKMGPSRAGTPEGKDRGNPGPVGLPSPRRPMSNL 965
QY 185 PWERDGAHSP-----EPKENVLPGDPTSNATSRGAG-PLPP----- 222
DB 966 WLKGDKSGSGSAGSNPGPRGDKGEAGRPGLPGALPGALIKVSGKRGPPGMGI 1025
QY 223 -----PSNPVAGAAG---GL--ALLILGVAAGGAMCMRRRAAPSE 260
DB 1026 RGLPGLKSSSGITGPPGMPGSGSGIGRSPGLPQASGLPKLKDNQTV--EISGSPGP 1083
QY 261 SRHPGPGSF--GRGSLG--GGGGGMPREAPRELIGIALRGGAADPPCPHYEVSQDVG 319
DB 1084 KGQPESEGFKTKGRDGLIGNIGPFGKKGEDGKTVS---GDVGLPGAPGFPVAGMRG 1139
QY 320 HP 321
DB 1140 EP 1141

RESULT 49
Q8TEJ5 PRELIMINARY; PRT; 705 AA.
ID Q8TEJ5;
AC Q8TEJ5;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE F1000201 protein (Fragment).
GN F1000201.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RY [1]
RS SEQUENCE FROM N.A.
RC TISSUE=SPLEEN;
RA Ukiyama H., Takano J., Nomura N., Kikuno R., Nagase T., Ohara O.;
RT "The nucleotide sequence of a long cDNA clone isolated from human
Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

```

```

DR EMBL; AK074129; BAB84955.1; -.
FT NON TER 1
SQ SEQUENCE 705 AA; 67430 MW; 07DB85A65A48ED3 CRC64;

Query Match 7.5%; Score 139; DB 4; Length 705;
Best Local Similarity 31.4%; Pred. No. 0.014;
Matches 66; Conservative 9; Mismatches 81; Indels 54; Gaps 11;

QY 159 RGKVLRLRVGQ-----SPRGAVPRKPVSEMERDRGAHSLRPGKENTLPGDPTSNATS 213
DB 204 RGLKMGNGVGGQGLPGAPGQGAAPGP--GLP-----GPAIGKXGLDGLPAPDPDKES 256
QY 214 -----RGAGPLPPEMPAVA--GAAGIALILGVAGGAGCAMRRRAKP-- 258
DB 257 GPPGVGPRGEGAVGPKPPGVVDGVVGAAG-----LPFGQSGSAGKGEPTGPPGL 311
QY 259 -----SESRHGP--GSFGRGSLGIGGGGGMPPREAPRELIGIALRGGAADPPCPHY 311
DB 312 IGPTGGMGLPGPKODRPAAPVPLGLDRGEPGDEPBGEPGGLG-----PFLGLGS 367
QY 312 EKVSQDYGHPIVVDGP-----PQSPNT 336
DB 368 AGLPGRGPP-----GPKGEAGPGGPGPV 391

RESULT 50
Q9D2V4 PRELIMINARY; PRT; 744 AA.
ID Q9D2V4;
AC Q9D2V4;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Procollagen, type VIII, alpha 1.
GN COL8A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RY [1]
RS SEQUENCE FROM N.A.
RC STRIN=ESTBL/6U; TISSUE=KIDNEY;
RX MEDLINE=21085660; Pubmed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukumishi Y., Komoto H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gasteierland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Mateno Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kanliya M., Lee N.H.,
RA Lyons P., Marchionni L., Washima J., Wazarelli J., Wombers P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Stoch K.F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK018742; BAB31383.1; -.
DR MGI; MGI:88463; Col8a1.
DR InterPro; IPR001073; C1q.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF00386; C1q; 1.
DR Pfam; PF01391; Collagen; 7.
DR PRINTS; PR00007; COMPLENCT1Q.
DR SMART; SM00110; C1Q; 1.
DR PROSITE; PS01113; C1Q; 1.
SQ SEQUENCE 744 AA; 73581 MW; C659BDCBCEDEB9C CRC64;

```

Query Match 7.5%; Score 139; DB 11; Length 744;  
Best Local Similarity 23.2%; Pred.No.0.014;  
Matches 83; Conservative 28; Mismatches 104; Indels 142; Gaps 18;

QY 67 PPGPHSSPNYEFYKYLIVG-----GAQGR-CEAPPAPNLLTCDRDPDLRFT 114  
|||  
Db 199 PGPPhGLPG-----IGKPGGPGLPQDPAKGERGPKGPQP----- 234  
|||  
QY 115 IKQEQYSPNLMGHEFRSHDYIIATSDGTR---EGLESLOG--GVCLTRGMKVLRYG 168  
|||  
Db 235 -----PGLQGPK-----GKKGMPGLPGLKPPGMHGP GPVGLPGVG 273  
|||  
QY 169 Q-----SPRGAVPRKPVSEMPMERDRGAHSLERKENTLPDP--TSN 210  
|||  
Db 274 KPGVTGFPGRQGLGKPGPRGERGPGGLIGVPGVGPFGMAGVGKPGQDGTFGQGFPGG 333  
|||  
QY 211 ATSRGABGLPPSPMPAVA-----GAAGLALL-----IGVAGAGAMCMR 252  
|||  
Db 334 KGEQGLPGLPGPGLPGVGKPGFPKGDRIIGVPGVLGPRGEKPIGAPGMGQ----- 388  
|||  
QY 253 RRRAKPSSESRHPG-PGSPGRGSLGL---GGGGGWMGPR---BAEPGELGI----- 295  
|||  
Db 389 -----PGEFGGLPGIPGMGPFGAIGFPGKGEQGVVFGPPGPKGEPGLQGFPGKPGFL 444  
|||  
QY 296 -----ALRG-----GGAADPPFCPHYEKVSGDYGHVPYIVODGPPQSPPI 336  
|||  
Db 445 GEVGPMPMRGLPGPIGPKGEGHKGLPGLPGVPGLLGPKGEP-GIPGDQGLQGPPI 500  
|||

Search completed: February 11, 2003, 12:05:05  
Job time : 36.2327 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: February 11, 2003, 11:47:24 ; Search time 14.5409 Seconds  
(without alignments)  
969.814 Million cell updates/sec

Title: US-10-021-121-4

Perfect score: 1850

Sequence: 1 MGPPHSGPGVAVGALLILG.....PVIIVDGPFGSPNNIYKV 340

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476326 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database : SwissProt\_40:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1850	100.0	340	EFB3_HUMAN	Q15768 homo sapien
2	1780	96.2	340	EFB3_MOUSE	O35393 mus musculu
3	637.5	34.5	336	EFB2_MOUSE	P52800 mus musculu
4	632	34.2	346	EFB1_HUMAN	P98172 homo sapien
5	629.5	34.0	333	EFB2_HUMAN	P52799 homo sapien
6	628	33.9	334	EFB1_CHICK	O7612 gallus galli
7	626.5	33.9	332	EFB2_BRARE	O73874 brachydacti
8	613.5	33.2	345	EFB1_MOUSE	P52795 mus musculu
9	608.5	32.9	345	EFB1_RAT	P52796 rattus norv
10	600	32.4	327	EFB1_XENLA	O13097 xenopus lae
11	185	10.0	195	EFB2_BRARE	P79727 brachydacti
12	179	9.7	238	EFB3_HUMAN	P52797 homo sapien
13	176	9.5	209	EFB2_MOUSE	P52801 mus musculu
14	175.5	9.5	213	EFB2_HUMAN	O43921 homo sapien
15	172	9.3	200	EFB2_CHICK	P52802 gallus galli
16	170.5	9.2	228	EFB5_CHICK	P52804 gallus galli
17	169.5	9.2	201	EFB4_HUMAN	P52798 homo sapien
18	169.5	9.2	228	EFB5_BRARE	P79728 brachydacti
19	167.5	9.1	216	EFB1_XENLA	P52794 xenopus lae
20	167.5	9.1	228	EFB5_HUMAN	P52803 homo sapien
21	167.5	9.1	228	EFB5_MOUSE	O08543 mus musculu
22	167.5	9.1	228	EFB5_RAT	P97605 rattus norv
23	166	9.0	205	EFB1_HUMAN	P20827 homo sapien
24	162.5	8.8	205	EFB1_RAT	P97553 rattus norv
25	161	8.7	205	EFB1_MOUSE	P52793 mus musculu
26	160.5	8.7	206	EFB4_MOUSE	O08542 mus musculu
27	159	8.6	680	EFB4_MOUSE	O08542 mus musculu
28	154.5	8.4	1049	CA13_BOVIN	P04258 bos taurus
29	151.5	8.2	301	CC02_CAEEL	P16565 caenorhabdi
30	148.5	8.0	1670	CA14_HUMAN	O01955 homo sapien
31	146.5	7.9	1527	CA1H_MOUSE	P13065 mus musculu
32	146	7.9	1027	CA1F_RIFPA	P30754 riftia pach
33	145	7.8	674	CA1A_BOVIN	P23206 bos taurus

## ALIGNMENTS

```

RESULT 1
EFB3_HUMAN
ID Q15768; O06680; Q92875; STANDARD; PRT; 340 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ephrin-B3 precursor (EPH-related receptor tyrosine kinase ligand 8)
DE (EPRK-8) (EPH-related receptor transmembrane ligand EPRK-13).
GN EPHB3 OR EPHB8 OR LERK8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
ON NCBI_Taxid=9606;
RX [1]
RP SEQUENCE FROM N.A.
RA Cerretti D.P.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96404527; PubMed=9126477;
RX MEDLINE=97271551; PubMed=9126477;
RA Tang X.X., Pleasure D.E., Ikegaki N.;
RT "cDNA cloning, chromosomal localization, and expression pattern of
RT EPHB8, a new member of the EPLG gene family encoding ligands of EPH-
RT related protein-tyrosine kinase receptors.";
RL Genomics 41:17-24(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain cortex;
RX MEDLINE=96404527; PubMed=8808709;
RA Gale N.W., Flenikien A., Compton D.C., Jenkins N.A., Copeland N.G.,
RA Galibert D.J., Davis S., Wilkinson D.G., Yancopoulos G.D.;
RT "Slk-13, a novel transmembrane ligand for the Eph family of receptor
RT tyrosine kinases, expressed in embryonic floor plate, roof plate and
RT hindbrain segments.";
RL Oncogene 13:1343-1352(1996).
RN [4]
RP FUNCTION: MAY PLAY A PIVOTAL ROLE IN FOREBRAIN FUNCTION. BINDS TO,
RP AND INDUCE THE COLLAPSE OF, COMMISSURAL AXONS/GROWTH CONES IN
RP VITRO. MAY PLAY A ROLE IN CONSTRAINING THE ORIENTATION OF
RP LONGITUDINALLY PROJECTING AXONS (BY SIMILARITY).
RP SUBCELLULAR LOCATION: Type I membrane protein.
RP TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BRAIN; EXPRESSED IN
RP EMBRYONIC FLOOR PLATE, ROOF PLATE AND HINDRAIN SEGMENTS.
RP -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U57001; AAC05170.1; -  
 DR EMBL; U66406; AAC51203.1; -  
 DR EMBL; U62725; AAC50707.1; -  
 DR Genew; HGNC:3228; EPNB3.  
 DR MIM; 602297; -  
 DR InterPro; IPR001799; Ephrin.  
 DR Pfam; PF00812; Ephrin; 1.  
 DR PRINTS; PR01347; EPHRIN.  
 DR PRODOM; PD002533; Ephrin; 1.  
 DR PROSITE; PS01299; EPHRIN; 1.  
 KM Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;  
 FM Signal; Polymorphism.  
 FT SIGNAL 1 27  
 FT CHAIN 28 340  
 FT DOMAIN 28 226  
 FT TRANSMEM 227 247  
 FT DOMAIN 248 340  
 FT DOMAIN 338 340  
 FT CARBOHYD 210 210  
 FT VARIANT 166 166  
 FT  
 SQ SEQUENCE 340 AA; 35834 MW; EDFF2A23C2FDE79F CRC64;  
 Query Match 100.0%; Score 1850; DB 1; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 3.7e-113;  
 Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPSPHSGGVAVGALLLLGLVGLVSGLSLEPVYNSANKRFOAGGYLYPQIDRLDL 60  
 DB 1 MGPSPHSGGVAVGALLLLGLVGLVSGLSLEPVYNSANKRFOAGGYLYPQIDRLDL 60  
 QY 61 LCPRAAPPSPHSPVYEFKYLIVGAGRCRCEAPPAENLLTCDRPLDLFTTKFOEY 120  
 DB 61 LCPRAAPPSPHSPVYEFKYLIVGAGRCRCEAPPAENLLTCDRPLDLFTTKFOEY 120  
 QY 121 SPNLWGHFRSHHDYIIATSDGTREGLESLOGGVCLTRGMKVLIRVQSPRGAVPRKP 180  
 DB 121 SPNLWGHFRSHHDYIIATSDGTREGLESLOGGVCLTRGMKVLIRVQSPRGAVPRKP 180  
 QY 181 VSEMERDRGAHSLPEKENTLPDPTSNATSRGAEGPLPSPMPAVAGAAGLALLL 240  
 DB 181 VSEMERDRGAHSLPEKENTLPDPTSNATSRGAEGPLPSPMPAVAGAAGLALLL 240  
 QY 241 GYAAGAGMCMRRRAKPSERHPGPGSFGGSGISLGCGGGMGPRAEPGLGIALRG 300  
 DB 241 GYAAGAGMCMRRRAKPSERHPGPGSFGGSGISLGCGGGMGPRAEPGLGIALRG 300  
 QY 301 GAADPPFCPHYEKVSGDYGHPIYIVODGPPSPPIYYKV 340  
 DB 301 GAADPPFCPHYEKVSGDYGHPIYIVODGPPSPPIYYKV 340

RESULT 2  
 EFB3 MOUSE STANDARD; PRT; 340 AA.  
 ID EFB3 MOUSE  
 AC 03533;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Ephrin-B3 precursor.  
 GN EPNB3.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCB1\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=98143367; PubMed=9484836;  
 RA Bergemann A.D., Zhang L., Chiang M.-K., Brambilla R., Klein R.,

RA Flanagan J.G.;  
 RT "Ephrin-B3, a ligand for the receptor EphB3, expressed at the midline  
 of the developing neural tube."  
 RL Oncogene 16:471-480(1998).

RP FUNCTION.  
 RX MEDLINE=20171264; PubMed=10704386;  
 RA Mondini R., Wideman C., Kaprielian Z.;  
 RT "Complementary expression of transmembrane ephrins and their receptors  
 in the mouse spinal cord: a possible role in constraining the  
 orientation of longitudinally projecting axons."  
 RL Development 127:1397-1410(2000).  
 CC -1- FUNCTION: MAY PLAY A PIVOTAL ROLE IN FOREBRAIN FUNCTION. BINDS TO,  
 AND INDUCE THE COLLAPSE OF, COMMISSURAL AXONS/GROWTH CONES IN  
 VITRO. MAY PLAY A ROLE IN CONSTRAINING THE ORIENTATION OF  
 LONGITUDINALLY PROJECTING AXONS.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED ON LATERAL FLOOR PLATE CELLS,  
 SPECIFICALLY ON COMMISSURAL AXON SEGMENTS THAT HAVE PASSED THROUGH  
 THE FLOOR PLATE. EXPRESSED IN CELLS OF THE RETINAL GANGLION CELL  
 LAYER DURING RETINAL AXON GUIDANCE IN THE OPTIC DISC.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE FLOOR PLATE THROUGHOUT THE  
 PERIOD OF COMMISSURAL AXON PATHFINDING.  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF025288; AAC3537.1; -  
 CC MGD; MGI:109196; Elnb3.  
 DR InterPro; IPR001799; Ephrin.  
 DR Pfam; PF00812; Ephrin; 1.  
 DR PRINTS; PR01347; EPHRIN.  
 DR PRODOM; PD002533; Ephrin; 1.  
 DR PROSITE; PS01299; EPHRIN; 1.  
 KM Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;  
 FM Signal.  
 FT SIGNAL 1 27  
 FT CHAIN 28 340  
 FT DOMAIN 28 227  
 FT TRANSMEM 228 248  
 FT DOMAIN 248 340  
 FT DOMAIN 338 340  
 FT CARBOHYD 210 210  
 FT  
 SQ SEQUENCE 340 AA; 35884 MW; 52F3D58FD209A6B8 CRC64;

Query Match 96.2%; Score 1780; DB 1; Length 340;  
 Best Local Similarity 95.6%; Pred. No. 1.2e-108;  
 Matches 325; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 MGPSPHSGGVAVGALLLLGLVGLVSGLSLEPVYNSANKRFOAGGYLYPQIDRLDL 60  
 DB 1 MGPSPHSGGVAVGALLLLGLVGLVSGLSLEPVYNSANKRFOAGGYLYPQIDRLDL 60  
 QY 61 LCPRAAPPSPHSPVYEFKYLIVGAGRCRCEAPPAENLLTCDRPLDLFTTKFOEY 120  
 DB 61 LCPRAAPPSPHSPVYEFKYLIVGAGRCRCEAPPAENLLTCDRPLDLFTTKFOEY 120  
 QY 121 SPNLWGHFRSHHDYIIATSDGTREGLESLOGGVCLTRGMKVLIRVQSPRGAVPRKP 180  
 DB 121 SPNLWGHFRSHHDYIIATSDGTREGLESLOGGVCLTRGMKVLIRVQSPRGAVPRKP 180  
 QY 181 VSEMERDRGAHSLPEKENTLPDPTSNATSRGAEGPLPSPMPAVAGAAGLALLL 240  
 DB 181 VSEMERDRGAHSLPEKENTLPDPTSNATSRGAEGPLPSPMPAVAGAAGLALLL 240  
 QY 241 GYAAGAGMCMRRRAKPSERHPGPGSFGGSGISLGCGGGMGPRAEPGLGIALRG 300  
 DB 241 GYAAGAGMCMRRRAKPSERHPGPGSFGGSGISLGCGGGMGPRAEPGLGIALRG 300

DB 241 GVAGAGCAGCRRRRAXPSESRHGPSPFRGSGSLGLGGGGGNGPREAREGELGIALRG 300

QY 301 GAADPPFCHEHYEKVSGDYGHPIYIVODGPPQSPFNIIYKV 340

DB 301 GTADPPFCHEHYEKVSGDYGHPIYIVODGPPQSPFNIIYKV 340

RESULT 3

EFB2\_MOUSE STANDARD; PRT; 336 AA.

ID EFB2\_MOUSE

AC P52800;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Ephrin-B2 precursor (EPH-related receptor tyrosine kinase ligand 5)

DE (LEK-5) (HTK ligand) (HTK-L) (ELF-2)

GN EPNB2 OR EPLG5 OR LERK5 OR HTKL OR ELF2 OR EPL5.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_Taxid=10090;

PN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96145238; PubMed=8559144;

RA Cerretti D.P., Vanden Bos T., Nelson N., Kozlosky C.J., Reddy P., Marschovsky E., Park L.S., Lyman S.D., Copeland N.G., Gilbert D.J., Jenkins N.A., Fletcher R.A.;

RA "Isolation of LERK-5: a ligand of the eph-related receptor tyrosine kinases";

RT Mol. Immunol. 32:1197-1205(1995).

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CB57BL/6J X SJL/J;

RX MEDLINE=95199254; PubMed=7534404;

RA Bennett B.D., Zeigler F.C., Gu Q., Fendly B., Goddard A.D., Gillett N., Matthews W.;

RA "Molecular cloning of a ligand for the EPH-related receptor protein-tyrosine kinase Hck";

RT Proc. Natl. Acad. Sci. U.S.A. 92:1866-1870(1995).

RL [3]

RP SEQUENCE FROM N.A.

RC STRAIN=ICR; TISSUE=Brain;

RX MEDLINE=95379837; PubMed=7651410;

RA Bergemann A.D., Cheng H.J., Brambilla R., Klein R., Flanagan J.G.;

RA "ELF-2, a new member of the Eph ligand family, is segmentally expressed in mouse embryos in the region of the hindbrain and newly forming somites";

RT Mol. Cell. Biol. 15:4921-4929(1995).

RL [4]

RP FUNCTION.

RX MEDLINE=20171264; PubMed=10704386;

RA Imondi R., Wideman C., Kaprielian Z.;

RA "Complementary expression of transmembrane ephrins and their receptors in the mouse spinal cord: a possible role in constraining the orientation of longitudinally projecting axons";

RT Development 127:1397-1410(2000).

RL [5]

CC -1- FUNCTION: MAY PLAY A ROLE IN CONSTRAINING THE ORIENTATION OF LONGITUDINALLY PROJECTING AXONS.

CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASE EPHB4.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- TISSUE SPECIFICITY: EXPRESSED ON LATERAL FLOOR PLATE CELLS, SPECIFICALLY ON COMMISSURAL AXON SEGMENTS THAT HAVE PASSED THROUGH THE FLOOR PLATE. EXPRESSED IN CELLS OF THE RETINAL GANGLION CELL LAYER DURING RETINAL AXON GUIDANCE TO THE OPTIC DISC.

CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE FLOOR PLATE THROUGHOUT THE PERIOD OF COMMISSURAL AXON PATHFINDING.

CC -1- PTM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE CYTOPLASMIC DOMAIN (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U16819; AAA99708.1; -

DR EMBL; L38847; AAC42052.1; -

DR EMBL; U30244; AAA82934.1; -

DR MGD; MG1:105097; Efnb2.

DR InterPro; IPR001799; Ephrin.

DR Pfam; PF00812; Ephrin.1.

DR PRINTS; PR01347; EPHRIN.

DR ProDom; PD002533; Ephrin.1.

DR PROSITE; PS01239; EPHRIN.1.

KW Developmental protein; Neurogenesis; Transmembrane; Glycoprotein; Signal; Phosphorylation.

FT SIGNAL 1 28

FT CHAIN 1 29

FT DOMAIN 29 336

FT TRANSMEM 233 253

FT DOMAIN 254 336

FT DOMAIN 334 336

FT CARBOHYD 339 339

FT CARBOHYD 142 142

FT CONFLICT 3 4

FT CONFLICT 177 177

SQ SEQUENCE 336 AA; 37202 MW; D0889496E39554 CRC64;

Query Match 34.5%; Score 637.5; DR 1; Length 336;

Best Local Similarity 42.1%; Pred. No. 1,1e-34;

Matches 139; Conservative 49; Mismatches 129; Indels 13; Gaps 5;

QY 14 GALLLVGLVGLVSGLSLEPYNNANRRFOAEGGYLVLPQIGRLDLCPRAPPGPHSS 73

DB 17 GLAVLCKRTAISRSIVALEPIYNNSSNEKFLPGQGLVLYPGIGKLDICPKV--DSKTV 73

QY 74 PNYEFLYLVGAGQGRCAAPPAHLLCDPDLRTFTKFOEYSPMLMGEPRSH 133

DB 74 GQYEVYVYVWDKQADRCITKENTPLNLCARPDDQVKTFIFOESFVLWGIFQNK 133

QY 134 DYIITSDGTRREGLESLGGVCLTRGMKVLIRVQ--SPRGAVPPKPPSEMPMER-DR 190

DB 134 DYIITSDGTRREGLESLGGVCLTRGMKVLIRVQ--SPRGAVPPKPPSEMPMER-DR 190

QY 191 GAHSLEPGKENTPGDPTSNATSRGAEGPIPPSPMAVAGAGLALLIGVAGAGAMC 250

DB 194 GRSSITPPFYKPNPSSITDNSAGHSNNLIGSEVALFAGIASGCIFIVYIITLVVLL 253

QY 251 WRRRRAPSESRHGPSPFRGSGSLGLGGGGGNGPREAREGELGIALRGGAADPPFCPH 310

DB 254 KYRRRRHKSPPQHTTTLSTLTATPKRGNN---NGSBSVDVILPLR---TADSVCPH 306

QY 311 YEKVSGDYGHPIYIVODGPPQSPFNIIYKV 340

DB 307 YEKVSGDYGHPIYIVODGPPQSPFNIIYKV 336

RESULT 4

EFB1\_HUMAN STANDARD; PRT; 346 AA.

ID EFB1\_HUMAN

AC P98172;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Ephrin-B1 precursor (EPH-related receptor tyrosine kinase ligand 2)

DE (LEK-2) (ELK ligand) (ELK-L).

GN EPNB1 OR EPLG2 OR LERK2 OR EPL-3.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI\_Taxid=9606;

PN [1]

RP SEQUENCE FROM N.A.

```

RC TISSUE=Placenta; PubMed=8070404;
RX Beckmann M.P., Czeretli D.P., Baum P., Vandenbos T., James L.,
RA Farrah T., Kozlosky C., Hollingsworth T., Shilling H., Maraskovsky E.,
RA Fletcher F.A., Lhotek V., Pawson T., Lyman S.D.;
RT "Molecular characterization of a family of ligands for eph-related
RT tyrosine kinase receptors."
RL EMBO J. 13:3757-3762(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Davis S., Gale N.W., Aldrich T.H., Maisompierre P.C., Lhotek V.,
RA Pawson T., Goldfarb W., Yancopoulos G.D.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Fletcher F.A., Huebner K., Shaffer L.G., Monaco A., Mueller U.,
RA Kozlosky C., Druck T., Simoneaux D.K., Fairweather N., Chelly U.,
RA Czeretli D.P., Belmont J.W., Beckmann M.P., Lyman S.D.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Howden P.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: BINDS TO, AND INDUCE THE COLLAPSE OF, COMMISSURAL
CC AXONS/GROWTH CONES IN VITRO. MAY PLAY A ROLE IN CONSTRAINING THE
CC ORIENTATION OF LONGITUDINALLY PROJECTING AXONS (BY SIMILARITY).
CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHB1 AND EPHA1.
CC BINDS GRIP1 AND GRIP2.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: HEART, PLACENTA, LUNG, LIVER, SKELETAL MUSCLE,
CC KIDNEY, PANCREAS.
CC -1- INDUCTION: BY TNF-ALPHA.
CC -1- PM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE
CC CYTOPLASMIC DOMAIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL; U09304; AAAS3093.1; -
DR EMBL; U37361; AAAS2369.1; -
DR EMBL; U09303; AAB41127.1; -
DR EMBL; AL136032; CAB86409.1; -
DR Genew; HGNC:3226; EFNB1.
DR MIM; 300035; -
DR InterPro; IPR001799; Ephrin.
DR Pfam; PF00812; Ephrin; 1.
DR PRINTS; PRO1347; EPHRIN.
DR PRODOM; PD002533; Ephrin; 1.
DR PROSITE; PS01299; EPHRIN; 1.
KW Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;
KW Signal; Phosphorylation.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 346 EPHRIN-B1.
FT DOMAIN 25 237 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 238 258 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 259 346 POTENTIAL.
FT DOMAIN 344 346 PDZ RECOGNITION MOTIF (POTENTIAL).
FT CARBOHYD 139 139 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 346 AA; 38006 MW; 473D2F1A5B89DE CRC64;

Query Match 34.2%; Score 632; DB 1; Length 346;
Best Local Similarity 39.5%; Pred. No. 2.5e-34;
Matches 145; Conservative 48; Mismatches 116; Indels 58; Gaps 9;

```

```

CY 8 PGQVAVGALLLGVGLVSGV-----SLPEYVNSANKRQAGGVYVQIGDRDL 61
4 PGQRWLGKVLAVMVWALCRLATPLAKNDLSEVWSLSNPKFLSGKGLVYIFKIDKDI 63

```

```

QY 62 CPARAPPGHSSPNYEFYLYVGAGQRCAPAPNLLTCDSPDIDRTIKOYS 121
DB 64 CPAAEGRP-----YLYLVVVRPEQAAACSTVDPNVLVCNPEQIRITIKOYS 118
QY 122 PNLWGHFERSHHDDYIATSDGTREGSLQGVCLTRGMKYLRLVGQS PRGAVPRKV 181
DB 119 PNMGLEPFKKHHDIYITSTNSNLSGLENRREGVCRTMTKIMKVGDDPNVATEQLTT 178
QY 182 SEMPERDRGAHSLR-PKENLPDGPITSNATSRAGCEPLPPSPAPVAAAGGLA---- 236
DB 179 SRPSKADVTVCATQAPESRSLSDSKKHTVNOEKSGP-----GASGGSSGDPD 231
QY 237 -----LLLVGAAGGA-----MCFRRRAKPSRSHHPGSPFGSGSLGL 277
DB 232 GPFNSKVALFAVAGACVIFLLIIFLYVLTLKRRKRHTQO-----PAAALSL 282
QY 278 ---GGGGMGSRERAPGELGIALRGGAADPPCFPHYKESGDYGVYIVQDGPPOS 333
DB 283 STLASPGGSGTAGTEPSDIIPLR---TTEVNYCPHYEKVSGDYGVYIVQEMPPOS 339
QY 334 PNLYKY 340
DB 340 ANIYKY 346

RESULT 5
ID EFNB2 HUMAN STANDARD; PRT; 333 AA.
AC P52759;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ephrin-B2 precursor (EPH-related receptor tyrosine kinase ligand 5)
DE (LEKRS-5) (HTK ligand) (HTK-L).
GN EFNB2 OR BR15 OR LERKS OR HTKL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96145238; PubMed=8559144;
RA Czeretli D.P., Vanden Bos T., Nelson N., Kozlosky C.J., Reddy P.,
RA Maraskovsky E., Park U.S., Lyman S.D., Copeland N.G., Gilbert D.J.,
RA Jenkins N.A., Fletcher R.A.;
RT "Isolation of LERK-5, a ligand of the eph-related receptor tyrosine
RT kinases."
RL Mol. Immunol. 32:1197-1205(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95199254; PubMed=7534404;
RA Bennett B.D., Zeigler F.C., Gu Q., Fendly B., Goddard A.D.,
RA Gillett N., Matthews W.;
RT "Molecular cloning of a ligand for the EPH-related receptor protein-
RT tyrosine kinase Htk."
RL Proc. Natl. Acad. Sci. U.S.A. 92:1866-1870(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=98192220; PubMed=9533549;
RA Vogt T., Stolz W., Welsh J., Jung B., Kerbel R.S., Kobayashi H.,
RA Landthaler M., McClelland M.;
RT "Overexpression of Lerk-5/Ep15 messenger RNA: a novel marker for
RT increased tumorigenicity and metastatic potential in human malignant
RT melanomas."
RL Clin. Cancer Res. 4:791-797(1998).
CC -1- FUNCTION: MAY PLAY A ROLE IN CONSTRAINING THE ORIENTATION OF
CC LONGITUDINALLY PROJECTING AXONS (BY SIMILARITY).
CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHB4 AND EPHA3.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: LONG AND KIDNEY.

```

```

CC -1- PTM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE
CC CYTOPLASMIC DOMAIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U16797; AAC9707.1; -
DR EMBL: U38734; AAC1752.1; -
DR EMBL: U81262; AAC03786.1; -
DR Genew; HGNC:3227; EPHN2.
DR MIM; 600527; -
DR InterPro; IPR001799; Ephrin.
DR Pfam; PF00812; Ephrin.1.
DR PRINTS; PR01347; EPHRIN.
DR ProDom; PD002533; Ephrin.1.
DR PROSITE; PS01299; Ephrin.1.
KW Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;
KW Signal; Phosphorylation.
FT CHAIN 1 27
FT SIGNAL 1 27
FT CHAIN 28 333
FT DOMAIN 28 229
FT TRANSMEM 230 250
FT DOMAIN 251 333
FT TRANSMEM 251 333
FT CARBOHYD 36 36
FT CARBOHYD 139 139
SQ SEQUENCE 333 AA; 36923 MW; 6D9332A632626A6A CRC64;

Query Match 34.0%; Score 629.5; DB 1; Length 333;
Best Local Similarity 40.9%; Pred. No. 3.5e-34;
Matches 135; Conservative 52; Mismatches 130; Indels 13; Gaps 5;

QY 14 GALLUSGLVGLVGLSLEPYVNSANKRPOAEGVYLTPQIGRLDLCPRARPPGPHS 73
DB 14 GULMTCRLRAISKISYLEPYYNNSKFLPGQGLVLYPQIGDKLIDCPKV---DSKIV 70
QY 74 PNYEFYKLYLVGAGORCEAPPAENLLTCDRPLDLRFTTKFQEYSNLMGHSEFRSH 133
DB 71 GQYEYKVMVVDQADRCIKENPTPLNCAKPPDDIKFTIKQFSPNLMGLFQNK 130
QY 134 DYITITSGTRGLESLOGGVCLTRGMKYLVRGQ--SPRGAVPRKVSSEMPER-DR 190
DB 131 DYIITSTNGSLGDLNQGCVQTRAMKILMKVGDASASTRKCDPTRRPELAGTN 190
QY 191 GAHSLPEPKENLPDPTSNATSGAELPPSPMPAVGAAGLALLLVGAAGAMC 250
DB 191 GRSTTSPFKRPPGSSSTGNSAGHGNMILGSEVALFAGISGCIIFVITLVLL 250
QY 251 WRRRAKPSBSRPPGSGSTRGGLGLGGGKGPPEAPGELGALRGGAADPPFC 310
DB 251 KVRRRRKSPHPTTTLSTLATPRSGNN---NGSEPSDIILPLR---TADSVFC 303
QY 311 YEKVSGDYHPYIVODGPSPPNYYKY 340
DB 304 YEKVSGDYHPYIVODGPSPPNYYKY 333

```

```

CC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
CC Gallus.
CC NCBI_TaxId=9031;
CC [1]
CC SEQUENCE FROM N.A.
CC MEDLINE=9723524; PubMed=9070326;
CC Holash J.A., Soans C., Chong L.D., Shao H., Dixit V.M.,
CC Pasquale E.B.;
CC "Reciprocal expression of the Eph receptor Csk5 and its ligand(s) in
CC the early retina."
CC Dev. Biol. 182:256-269(1997).
CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASE EPHB2.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE
CC CYTOPLASMIC DOMAIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U72394; AAC07986.1; -
DR InterPro; IPR001799; Ephrin.
DR Pfam; PF00812; Ephrin.1.
DR PRINTS; PR01347; EPHRIN.
DR ProDom; PD002533; Ephrin.1.
DR PROSITE; PS01299; Ephrin.1.
KW Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;
KW Signal; Phosphorylation.
FT CHAIN 1 25
FT SIGNAL 1 25
FT CHAIN 26 334
FT DOMAIN 26 231
FT TRANSMEM 232 252
FT DOMAIN 253 334
FT DOMAIN 332 334
FT CARBOHYD 133 133
SQ SEQUENCE 334 AA; 36858 MW; 48AF56B9E56CD5 CRC64;

Query Match 33.9%; Score 628; DB 1; Length 334;
Best Local Similarity 39.7%; Pred. No. 4.4e-34;
Matches 146; Conservative 50; Mismatches 100; Indels 72; Gaps 13;

QY 8 PGYVR--VGALLIGVLVGLSLEPYVNSANKRPOAEGVYLTPQIGRLDLCPR 65
DB 4 PGRGKMLIGVLMLCRLAAPLAKSLFVSWAGNPKFMSGKGVIVYEIDCKLIDCPKA 63
QY 66 RPPGPHSPNYEFYKLYLVGAGORCEAPPAENLLTCDRPLDLRFTTKFQEYSNLM 125
DB 64 EPKRP-----YDYIKLYLVKKDQADACSTMDPNVLTCKRPPQGRIFRTTKFQ 118
QY 126 GHEFRSHDYIATSDGTREGLESLOGGVCLTRGMKYLVRGQSPRGAVPRKVS 185
DB 119 GLEFRQODFYITSTNGTLDGLENREGVCCQTRSKTIWKVQCDP-NAVIFQLT 177
QY 186 MER-----DGAHSL-----EPKKNLPDPTNA--TSRGAEGELPPMPA 231
DB 178 SKENDTVKIVTQSPKHKVTVBEPK--PGSVNONGCTQGPSDFL--SKVAVFA 232
QY 232 AGG-----LALLLVGAAGAMCRRRAKPSBSR-----PPGSGFRG 272
DB 233 IGAGCVIFILITLVLLIKI-----RKRHRKTKQRAALSLSTLAPKSGNA 283
QY 273 GSLGSGGGMGRPEAPGELGALRGGAADPPFCGHYKVSQDVGHPRYYIVODGP 332
DB 284 GS-----EPSDIILPLR--TTENNYGPHYKVSQDVGHPRYYIVODGP 326
QY 333 PPNYYKY 340

```



Db 327 PANIYKY 334

# RESULT 7

EFB2\_BRARE EFB2\_BRARE STANDARD; PRT; 312 AA.

ID EFB2\_BRARE

AC 073874

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE EFB2 BR precursor.

OS EFB2 OR EFB2A.

OC Brachydanio rerio (Zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

OC Cyprinidae; Danio.

OC NCBI\_Taxid=7955;

OX NCBI\_Taxid=7955;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98438455; PubMed=9765210;

RA Durbin L., Brennan C., Shiomi K., Cooke J., Barrios A.,

RA Shanmugasundaram S., Guthrie B., Lindberg R., Holder N.,

RT "Eph signaling is required for segmentation and differentiation of

the somites."

RL Genes Dev. 12:3096-3109(1998).

CC -1 SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASE EPHB4.

CC -1 SUBCELLULAR LOCATION: Type I membrane protein.

CC -1 PTM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE

CYTOPLASMIC DOMAIN (BY SIMILARITY).

CC -1 SIMILARITY: BELONGS TO THE EPHRIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation

at the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; AJ004863; CA006168.1; -

DR ZFIN; ZDB-GENE-990415-67; efnb2a.

DR InterPro; IPR001799; Ephrin.

DR Pfam; PF00812; Ephrin.1.

DR PRINTS; PR01347; EPHRIN.

DR PRODOM; PD002533; Ephrin.1.

DR PROSITE; PS01299; EPHRIN.1.

KW Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;

KW Signal; Phosphorylation.

FT SIGNAL 1 24

FT CHAIN 25 332

FT DOMAIN 25 225

FT TRANSMEM 226 246

FT DOMAIN 247 332

FT DOMAIN 330 332

SEQUENCE 332 AA; 36724 MW; 189ED8237C71C8B CRC64;

Query Match

Best Local Similarity 42.28; Pred. No. 5.4e-34; Length 332;

Matches 145; Conservative 54; Mismatches 106; Indels 39; Gaps 12;

QY 14 GALLLGLVGVSGSLSPVYNSANKRFOAEGGVLYPCIGDRLLDLPAPAPPGPHSS 73

Db 11 GVLVACVKNVSRALILDSITWNTNTKTFVPGQGLVLYPCIGDRLLDLPAPAPPGPHSS 67

QY 74 PNYEYKYLIVGAGRGCEAPAPNLLITCDPRDLIRFTIKQVSPNMGHEFFSHH 133

Db 68 EGVEYKLYMPPLKQKQVTKADTPILNCVKKPDQVKTFLKQEPSPNMGHEFFRGK 127

QY 134 DYYIATSDGTFEESLQSGVCLTRGKVLRLVQSPRGAGVPRK-PVSEMPERDGA 192

Db 128 DYYIATSDGTFEESLQSGVCLTRGKVLRLVQSPRGAGVPRK-PVSEMPERDGA 187

QY 193 AHS-----LEP-----GKENTPGDPTSNATSRGAGPLPPSPMAVAGAGGLALLLIGV 242

Db 186 KDSKSNVTLKPPASPHGEDK--GDGNKSSVIGSEVAL-----FACIASAIVIIIIIML 241

QY 243 AGAGGAMCWRRRARPPSSRRPSPGSGF-----RGSLGLGGGCGMPREAPPELGA 296

Db 242 VFL--LTKYRRHRHRS--PQHTTLSTLSTLTPKKGGS-----GGNNNG---SEPDIIIP 291

QY 297 LRGGAADPPPCPEHYEKVSGDGHFVYIVODGPPSPNIIYKYV 340

Db 292 LR---TADSVFCPEHYEKVSGDGHFVYIVODGPPSPNIIYKYV 332

# RESULT 8

EFB1\_MOUSE EFB1\_MOUSE STANDARD; PRT; 345 AA.

ID EFB1\_MOUSE

AC P52755;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Ephrin-B1 precursor (EPH-related receptor tyrosine kinase ligand 2)

DE (LEK-2) (ELK ligand) (ELK-1) (STRAL protein) (CEK5 receptor ligand)

DE (CEK5-L).

OS EPHB1 OR EPLG3 OR LEK2 OR STRAL OR EPL2.

OC Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI\_Taxid=10090;

OX NCBI\_Taxid=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=129/SV.

RX MEDLINE=95203867; PubMed=7896266;

RA Fletcher F.A., Renshaw B., Hollingsworth T., Baum P., Lyman S.D.,

RA Jenkins N.A., Gilbert D.J., Copeland N.G., Davidson B.L.,

RT "Genomic organization and chromosomal localization of mouse Eplg2, a

gene encoding a binding protein for the receptor tyrosine kinase

alk."

RL Genomics 24:127-132(1994).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=95377533; PubMed=7649373;

RA Boullier P., Oulad-Abdelghani M., Vicaire S., Garnier J.M.,

RA Schubbauer B., Dolle P., Chambon P.

RT "Efficient cloning of cDNAs of retinoic acid-responsive genes in Plg

embryonal carcinoma cells and characterization of a novel mouse gene,

Stral (mouse LEK-2/Eplg2)."

RT Dev. Biol. 170:420-433(1995).

RL [3]

RN SEQUENCE FROM N.A.

RP TISSUE=Brain;

RX MEDLINE=95014510; PubMed=7929389;

RA Shao H., Lou D., Pandey A., Pasquale E.B., Dixit V.M.,

RT "cDNA cloning and characterization of a ligand for the tyrosine kinase

protein-tyrosine kinase."

RT J. Biol. Chem. 269:26606-26609(1994).

RN [4]

RP FUNCTION.

RX MEDLINE=20171264; PubMed=10704386;

RA Imondi R., Wideman C., Kaprielian Z.

RT "Complementary expression of transmembrane ephrins and their receptors

in the mouse spinal cord: a possible role in constraining the

RT orientation of longitudinally projecting axons."

RL Development 127:1397-1410(2000).

CC -1 FUNCTION: BINDS TO, AND INDUCE THE COLLAPSE OF, COMMISSURAL

AXONS/GROWTH CONES IN VITRO. MAY PLAY A ROLE IN CONSTRAINING THE

ORIENTATION OF LONGITUDINALLY PROJECTING AXONS.

CC -1 SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHB1 AND EPHB2.

CC -1 TISSUE SPECIFICITY: EXPRESSED ON LATERAL FLOOR PLATE CELLS.

CC SPECIFICALLY ON COMMISSURAL AXON SEGMENTS THAT HAVE PASSED THROUGH

THE FLOOR PLATE. EXPRESSED IN CELLS OF THE RETINAL GANGLION CELL

CC LAYER DURING RETINAL AXON GUIDANCE TO THE OPTIC DISC.

CC -1 DEVELOPMENTAL STAGE: EXPRESSED IN THE FLOOR PLATE THROUGHOUT THE

CC PERIOD OF COMMISSURAL AXON PATHFINDING.

-1- PTM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE CYTOPLASMIC DOMAIN.  
 -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U07602; AAC53247.1; -  
 DR EMBL: U07598; AAC53247.1; JOINED.  
 DR EMBL: U07599; AAC53247.1; JOINED.  
 DR EMBL: U07600; AAC53247.1; JOINED.  
 DR EMBL: Z48781; CAA88695.1; -  
 DR EMBL: U12983; AAA53231.1; -  
 DR MGJ: MGJ102708; Efbn1.  
 DR InterPro: IPR001799; Ephrin.  
 DR Pfam: PF00812; Ephrin.1.  
 DR PRINTS: PR01347; EPHRIN.  
 DR ProDom: PD002533; Ephrin.1.  
 DR PROSITE: PS01299; EPHRIN.1.  
 KM Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;  
 KW Signal; Phosphorylation.  
 FT CHAIN 1 24  
 FT SIGNAL 1 24  
 FT DOMAIN 25 345  
 FT TRANSMEM 237 257  
 FT DOMAIN 237 257  
 FT DOMAIN 237 257  
 FT DOMAIN 237 257  
 FT CARBOHYD 139 139  
 FT CONFLICT 90 90  
 FT SEQUENCE 345 AA; 37859 MW; 8C96FD3DC5BC405 CRC64;  
 SQ  
 Query Match 33.2%; Score 613.5; DB 1; Length 345;  
 Best Local Similarity 38.2%; Pred. No. 3.9e-33;  
 Matches 138; Conservative 51; Mismatches 107; Indels 65; Gaps 10;

DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Ephrin-B1 precursor (EPH-related receptor tyrosine kinase ligand 2)  
 DE (LERK-2) (ELK ligand) (ELK-L).  
 OS EPHB1 OR EPHB2 OR LERK2.  
 GN Rattus norvegicus (Rat).  
 CC Eukaryote; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=95022634; PubMed=7936649;  
 RA Fletcher P.A., Carpenter M., Shilling H., Baum P., Ziegler S.,  
 RA Gimpel S., Hollingworth T., Vandenbos T., Davison B.L.,  
 RA Lyman S.D., Beckmann M.P.,  
 RT "LERK-2, a binding protein for the receptor-tyrosine kinase Elk, is  
 RT evolutionarily conserved and expressed in a developmentally regulated  
 RT pattern";  
 RL Oncogene 9:3241-3248(1994).  
 CC -1- FUNCTION: BINDS TO, AND INDUCE THE COLLAPSE OF, COMMISSURAL  
 CC AXONS/GROWTH CONES IN VITRO. MAY PLAY A ROLE IN CONSTRAINING THE  
 CC ORIENTATION OF LONGITUDINALLY PROJECTING AXONS (BY SIMILARITY).  
 CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHB3 (PREFERRED),  
 CC EPHB1 AND EPHB2.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- PTM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE  
 CC CYTOPLASMIC DOMAIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U07560; AAA53092.1; -  
 DR InterPro: IPR001799; Ephrin.  
 DR Pfam: PF00812; Ephrin.1.  
 DR PRINTS: PR01347; EPHRIN.  
 DR ProDom: PD002533; Ephrin.1.  
 DR PROSITE: PS01299; EPHRIN.1.  
 KM Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;  
 KW Signal; Phosphorylation.  
 FT CHAIN 1 24  
 FT SIGNAL 1 24  
 FT DOMAIN 25 345  
 FT TRANSMEM 237 257  
 FT DOMAIN 237 257  
 FT DOMAIN 237 257  
 FT CARBOHYD 139 139  
 FT SEQUENCE 345 AA; 37951 MW; 1B3045C5C73587E CRC64;  
 SQ  
 Query Match 32.9%; Score 608.5; DB 1; Length 345;  
 Best Local Similarity 38.0%; Pred. No. 8.2e-33;  
 Matches 137; Conservative 52; Mismatches 107; Indels 65; Gaps 10;

```

Db 189 VXT-----ATQPRGRSGSDSDGKETVNOQEKSGPAGGSGSSDTSDFNSK 236
QY 227 LLLLVAGAGG-----MCMRRRAKPSERHRHGPSPFGSGSLG-----CG 279
Db 237 VALFAAVGAGCVIFLLIIIFLVLLIKRKRRKHTQQ-----DAALSLSTLASP 287
QY 280 GGGMGPREAPGELGIALRGGAADPPFCPHYKVSQDGHVYIVQDGPSPENIYK 339
Db 288 KQDSGAGTEPEDIIIFLR---TENNVCPHYKVSQDGHVYIVQEMPSPANIYK 344
QY 340 V 340
Db 345 V 345

RESULT 10
EPFL_XENLA STANDARD; PRT; 327 AA.
ID EPFL_XENLA
AC 013097;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Ephrin-B1 precursor (EPH-related receptor tyrosine kinase ligand 2) (LERK-2) (Elk ligand) (ELK-L) (XLERK).
GN EPNBI OR EPLG2 OR LERK2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OC NCBI_Taxid=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97316777; PubMed=9174051;
RA Jones T.L., Karavanova I., Chong L., Zhou R.P., Daar I.O.;
RT "Identification of Xlerk, an Eph family ligand regulated during mesoderm induction and neurogenesis in Xenopus laevis.";
RL Oncogene 14:2159-2166(1997).
CC -1- FUNCTION: MAY HAVE A ROLE IN THE DEVELOPING MESENCHYMAL AND NERVOUS TISSUE.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: IN THE ADULT, EXPRESSED AT LOW LEVELS IN MOST ADULT TISSUES WITH INCREASED LEVELS OBSERVED IN THE KIDNEY, OOCYTES, OVARY AND TESTIS.
CC -1- PM: INDUCIBLE PHOSPHORYLATION OF TYROSINE RESIDUES IN THE CYTOPLASMIC DOMAIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch)..
CC EMBL; U31427; AAC35995.1; -.
CC InterPro: IPR001799; Ephrin.
CC Pfam: PF00812; Ephrin.1.
CC PRINTS; PR01347; EPHRIN.
CC ProDom; PD002533; Ephrin.1.
CC PROSITE; PS01999; EPHRIN.1.
KM Developmental protein; Neurogenesis; Transmembrane; Glycoprotein;
KM Signal; Phosphorylation.
FT CHAIN 1 20 POTENTIAL.
FT SIGNAL 1 20 EPHRIN-B1.
FT DOMAIN 21 225 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 226 246 POTENTIAL.
FT DOMAIN 247 327 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 328 340 PDZ RECOGNITION MOTIF (POTENTIAL).
FT CAROYND 131 131 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROYND 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 327 AA; 36621 MW; 71230CE7F6B5974 CRC64;

```

```

Query Match 32.4%; Score 600; DB 1; Length 327;
Best Local Similarity 39.8%; Pred. No. 2,8e-32;
Matches 146; Conservative 43; Mismatches 100; Indels 78; Gaps 12;

QY 10 GVR--VQALLIGVGLVSGLSLEPYWNSANKRFOAEGVLYIPIQIRDLICPRA-- 65
Db 3 GRRRLGLLVIVIRLCSAAGKNLEPYTWNSQNPFRISGLVLYEIGRDLIIICPKGF 62
QY 66 RPPGPHSPSEYEFYKLYLVGAQGRCEAPAPNLLTCDRPPDLIRFTIKFOEYSPNLM 125
Db 63 QP-----YBYKLYMRDQLEACSTVDPNVLYTNOQGEKRFYTIKQBSFNIM 114
QY 126 GHEFRSHDYIILATSDTREGLESLOGVCUTRGKVLRYVQSGPRGAVRKPYSKMP 185
Db 115 GLEFRNODYIYITNSTNSITLQLENREGVCTRSKIMTKVGDPP--NAVPEQLT--- 169
QY 186 MERDRGAASLSERKENLGDPTSNATSRGA-EGPLPP-----SMPAVAGAAGLA 236
Db 170 -----TTPRSKE--ADNTGKIATGPNNGPVQNGKSDTMSDKPTGRWGVDFE 217
QY 237 LLLLVAGAGGAGC-----MRRRAKPS-----SRHPGSGFGRG 273
Db 218 NSKIATFAAIGAGCVIFLLIIIFLVLLIKRKRRKHTQQAALSLSTLASPKSGNAG 277
QY 274 SLGGGGGGMGPREAPGELGIALRGGAADPPFCPHYKVSQDGHVYIVQDGPSP 333
Db 278 S-----EPSDIIFLR--TENNVCPHYKVSQDGHVYIVQEMPSP 320
QY 334 PNYYKY 340
Db 321 ANIYKY 327

RESULT 11
EPF2 BRARE STANDARD; PRT; 195 AA.
ID EPF2 BRARE
AC P79727;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Ephrin-A2 precursor (EPH-related receptor tyrosine kinase ligand 6) (LERK-6) (ELF-1) (ZEPHL3).
GN EPNA2 OR EPLG6 OR LERK6.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OC NCBI_Taxid=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97195707; PubMed=9043080;
RA Brennan C., Monschau B., Lindberg R., Guthrie B., Dreacher U.,
RA Bonhoeffer F., Holder N.;
RT "Two Eph receptor tyrosine kinase ligands control axon growth and may be involved in the creation of the retinotectal map in the zebrafish.";
RL Development 124:655-664(1997).
CC -1- FUNCTION: CONTROL AXON GROWTH AND MAY BE INVOLVED IN THE CREATION OF THE RETINO-TECTAL MAP.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (Potential).
CC -1- TISSUE SPECIFICITY: WIDESPREAD EXPRESSION IN THE EMBRYO.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE PRESUMPTIVE MIDBRAIN OF DEVELOPING EMBRYOS FROM THE SIX-SOMITE STAGE. BY 24 HOURS IT IS EXPRESSED THROUGHOUT THE MIDBRAIN INCLUDING THE REGION OF THE PRESUMPTIVE TECTUM. AT LATER STAGES IT IS EXPRESSED IN A GRADED FASHION THROUGHOUT THE TECTUM.
CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

```



(2)  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Brain;  
 RX MEDLINE=95181289; PubMed=7876076;  
 RA Shao H., Lou L., Pandey A., Verdier M.F., Siever D.A., Dixit V.K.;  
 RT "cDNA cloning and characterization of a Cdk7 receptor  
 RT protein-tyrosine kinase ligand that is identical to the ligand  
 RT (EphA-1) for the Mek-4 and Ser receptor protein-tyrosine kinases";  
 RL J. Biol. Chem. 270:3467-3470(1995).  
 CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA3, EPHA4 AND  
 CC EPHA5.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
 CC (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U14941; AAA53636.1; -  
 DR EMBL: U14752; AAA68520.1; -  
 DR MGD: MG1:102707; Etna2.  
 DR InterPro: IPR001799; Ephrin.  
 DR Pfam: PF00812; Ephrin; 1.  
 DR PRINTS: PRO1347; Ephrin; 1.  
 DR ProDom: PD002533; Ephrin; 1.  
 DR PROSITE: PS01299; Ephrin; 1.  
 KM Glycoprotein; GPI-anchor; Signal.  
 FT SIGNAL 1 20  
 FT CHAIN 21 209  
 FT CARBOHYD 38 38  
 FT CARBOHYD 170 170  
 FT CARBOHYD 184 184  
 SQ SEQUENCE 209 AA; 23586 MW; P197545F25B9ABC CR64;  
 Query Match 9.5%; Score 176; DB 1; Length 209;  
 Best Local Similarity 29.3%; Pred. No. 4.9e-05;  
 Matches 58; Conservative 19; Mismatches 69; Indels 52; Gaps 7;  
 QY 33 VYVNSANKRFOAE-----GGVLYPQIGDRDLCPARPGRPHSSNVEFYKLYLVGA 87  
 DB 35 VYVNSNRFRFOYSAVGGGGITVEVINDYLDYCPHYGALP-PARMERIILVNVGE 93  
 QY 88 QGRCEAPAPNLLITCDRPL--DNRFTIKFOEYSPNLMGHEFRSHDYIATSDGT 144  
 DB 94 GHASCDHRGQRFKMECNRPAPAGPLKFSKFLTFPFSLGFEFRGHEYYISATP-- 151  
 QY 145 REGESLOGGVCITRGMKYLIRVGSPPRGAVPRKPSSEMEMDRGAASLEPKENLP 204  
 DB 152 ---PNLVDRPCLR-LKYYVR-----PTNETLY 174  
 QY 205 GDP-----TSNATSRGAEG 218  
 DB 175 EAPERTFTSNSSCSGLG 192  
 RESULT 14  
 EFA2 HUMAN STANDARD; PRT; 213 AA.  
 AC 043921; 076020;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ephrin-A2 precursor (Eph-related receptor tyrosine kinase ligand 6)  
 DE (LRRK-6) (HEK7-ligand) (HEK7-L).  
 GN EFA2 OR EPLG6 OR LRRK6.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98126446; PubMed=9465306;  
 RA Cerretti D.P., Nelson N.;  
 RT "Characterization of the genes for mouse LRRK-3/Ephrin-A3 (Epl3)  
 RT mouse LRRK-4/Ephrin-A4 (Epl4), and human LRRK-6/Ephrin-A2 (Epl6);  
 RT conservation of intron/exon structure";  
 RL Genomics 47:131-135(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Lamerdin J.E., McCreedy P.W., Skowronski E., Adamson A.W.,  
 RA Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,  
 RA Phan H., Velasco N., Garsen J., Dangnan L., Poundstone P.,  
 RA Christensen M., Georgesou A., Avila J., Liu S., Attix C., Andreise T.,  
 RA Trakshim M., Amico-Keller G., Coefield J., Dharre S., Lucas S.,  
 RA Bruce R., Thomas P., Quan G., Krommiller B., Arellano A.,  
 RA Montgomey M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.O.,  
 RA Carraro A.V.;  
 RA Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=Brain;  
 RX MEDLINE=99045414; PubMed=9826538;  
 RA Aashem H.C., Pedetour F., Grosgeorge J., Logtenberg T.;  
 RT "Cloning, chromosomal mapping, and tissue expression of the gene  
 RT encoding the human Eph-family kinase ligand ephrin-A2";  
 RL Biochem. Biophys. Res. Commun. 252:378-382(1998).  
 CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA3, EPHA4 AND  
 CC EPHA5.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
 CC (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U92896; AAC39577.1; -  
 DR EMBL: U92893; AAC39577.1; JOINED.  
 DR EMBL: AC004286; AAC04836.1; -  
 DR EMBL: AJ007292; CA007435.1; -  
 DR GeneW; HGNC:3222; EPHA2.  
 DR MIM; 602756; -  
 DR InterPro: IPR001799; Ephrin.  
 DR Pfam: PF00812; Ephrin; 1.  
 DR PRINTS: PRO1347; Ephrin; 1.  
 DR ProDom: PD002533; Ephrin; 1.  
 DR PROSITE: PS01299; Ephrin; 1.  
 KM Glycoprotein; GPI-anchor; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 213  
 FT CARBOHYD 42 42  
 FT CARBOHYD 174 174  
 FT CARBOHYD 188 188  
 FT CARBOHYD 188 188  
 FT CONFLICT 6 6  
 FT CONFLICT 25 26  
 FT CONFLICT 29 30  
 FT CONFLICT 30 30  
 SQ SEQUENCE 213 AA; 23878 MW; 33C9FB1A8168B2D0 CR64;  
 Query Match 9.5%; Score 175.5; DB 1; Length 213;  
 Best Local Similarity 36.8%; Pred. No. 5.4e-05;  
 Matches 43; Conservative 14; Mismatches 51; Indels 9; Gaps 3;  
 QY 33 VYVNSANKRFOAE-----GGVLYPQIGDRDLCPARPGRPHSSNVEFYKLYLVGA 87  
 DB 39 VYVNSNRFRFOYSAVGGGGITVEVINDYLDYCPHYGALP-PARMERIILVNVGE 97

QY 88 QGRRCCEAPPANLLITCDRDL---DIRFTIKQEYSPNMGHFRSHDYTIATS 141  
DB 98 GHASCDHRORGFKEWECNRPAAPGPIKFSBKFLFTFPELGFEPFGHEYYIATS 154

## RESULT 15

EPF2\_CHECK STANDARD; PRT; 200 AA.

AC P52802;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Ephrin-A2 precursor (EPH-related receptor tyrosine kinase ligand 6)  
DE (LRRK-6) (ELF-1).  
GN EFN2 OR EPLG6 OR LRRK6 OR ELF1.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
NCBI\_TaxID=9031;

QY SEQUENCE FROM N.A.  
RX MEDLINE=95360981; PubMed=7634327;  
RA Cheng H.J., Nakamoto M., Bergmann A.D., Flanagan J.G.;  
RT "Complementary gradients in expression and binding of ELF-1 and Mek4  
in development of the topographic retinorectal projection map";  
RL Cell 82:371-381(1995).

CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA3, EPHA4 AND  
CC EPHAS (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
(Potential).

CC -1- TISSUE SPECIFICITY: EXPRESSED IN A GRADIENT ACROSS THE TECTUM  
CC BEING MORE STRONGLY EXPRESSED AT THE POSTERIOR POLE.

CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; I40933; AAC42229.1; -  
DR InterPro: IPR001799; Ephrin.  
DR Pfam: PF00812; Ephrin.1.  
DR PRINTS; PR01347; EPHRIN.  
DR ProDom; PD002533; Ephrin.1.  
DR PROSITE; PS01299; EPHRIN.1.  
KW Glycoprotein; GPI-anchor; Signal.  
FT SIGNAL 1 22  
FT CHAIN 1 22  
FT CARBOHYD 23 200  
FT CARBOHYD 161 161  
FT CARBOHYD 175 175  
SQ SEQUENCE 200 AA; 23049 MW; 8FAB1A5E5EED96 CRC64;

Query Match 9.3%; Score 172; DB 1; Length 200;

Best Local Similarity 35.3%; Pred. No. 8.5e-05; Indels 16; Gaps 5;

Matches 49; Conservative 16; Mismatches 58;

QY 15 ALLILGLVGLVSGISLEP-----VYNSANKRFOAGGVLYPQIGDRDLCPRA 65

DB 7 AALLIAIVG-VCVMSDDPGKVIDRYAVVYNNRGNPRFH-RGDTVEVSIINDYLDICPHY 64

QY 66 RPPGPHSPHYEYKLYVAGAGRCCEAPPANLLITCDRDL---DIRFTIKQEYSP 122

DB 65 EEPILP-AEMKEIVLWVYVEGHASCDHRQCKFKWECNRPDSPGPIKFSBKFLFTP 122

QY 123 NMGHEFRSHDYTIATS 141

DB 123 FSLGFEPFGHEYYIATS 141

RESULT 16  
EPF5\_CHECK STANDARD; PRT; 228 AA.

AC P52804;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Ephrin-A5 precursor (EPH-related receptor tyrosine kinase ligand 7)  
DE (LRRK-7) (RAGS protein).  
GN EPHAS OR RAGS.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
NCBI\_TaxID=9031;

QY SEQUENCE FROM N.A.  
RX MEDLINE=95360980; PubMed=7634326;  
RA Drescher U., Kremoser C., Handwerker C., Loschinger J., Noda M.,  
RA Bonhoeffer F.;  
RT "In vitro guidance of retinal ganglion cell axons by RAGS, a 25 kDa  
tectal protein related to ligands for Eph receptor tyrosine  
kinases";  
RL Cell 82:359-370(1995).

CC -1- FUNCTION: INDUCES GROWTH CONE COLLAPSE AND REPUSION OF RETINAL  
CC GANGLION CELL AXONS.

CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
(Potential).

CC -1- TISSUE SPECIFICITY: EXPRESSED IN A GRADED FASHION ACROSS THE  
CC TECTUM BEING MORE STRONGLY EXPRESSED TOWARDS THE POSTERIOR POLE.

CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X90377; CA62027.1; -  
DR InterPro: IPR001799; Ephrin.  
DR Pfam: PF00812; Ephrin.1.  
DR PRINTS; PR01347; EPHRIN.  
DR ProDom; PD002533; Ephrin.1.  
DR PROSITE; PS01299; EPHRIN.1.  
KW Developmental protein; Neurogenesis; Glycoprotein; GPI-anchor; Signal.  
FT SIGNAL 1 20  
FT CHAIN 1 228  
FT CARBOHYD 21 228  
FT CARBOHYD 37 37  
SQ SEQUENCE 228 AA; 26206 MW; 56DB84FBD8CF18AD CRC64;

Query Match 9.2%; Score 170.5; DB 1; Length 228;

Best Local Similarity 28.9%; Pred. No. 0.00012;

Matches 73; Conservative 30; Mismatches 91; Indels 59; Gaps 13;

QY 16 LLLIGVGLV-VSGISLEP-----VYNSANKRFOAGGVLYPQIGDRDLCPRA 64

DB 6 MLLIAVALLWCVRGQEPGRKAVADRYAVVYNNSTNPRFQ-QGDHIVDCINDYLDVCPH 64

QY 65 ARPGRPHSPHYEYKLYVAG-----GAQGRRCCEAPPANLLITCDRDLCPRA 112

DB 65 YEDVPEDEKT--ERYVLYVWVDFGYSCHISGFKWECNRPHSPV-----GPIK 113

QY 113 FTTFQEYSPHYEYKLYVAGAGRCCEAPPANLLITCDRDL---DIRFTIKQEYSP 169

DB 114 FSEKFLFTFPELGFEPFGHEYYIATSALPDNGRRS-----CLK-LKVFV--- 159

QY 170 SPRGAVPRKPVSEMPERDCAHSLPEKENLPGDPTSNATSRGAGCPPLPPSNP--A 227

DB 160 -PANSCKMTIGVHDVRVVDVNDKVENSLPADDTV---RESAEPGRG-ENNAQTPRIPIRL 214

QY 228 VAGAAGLALLL 240  
 DB 215 LATLLFLAMLLI 227

RESULT 17  
 EF44\_HUMAN STANDARD; PRT; 201 AA.

AC P52798; O95457;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ephrin-A4 precursor (EPH-related receptor tyrosine kinase ligand 4)  
 DE (LEK-4)  
 GN EPRNA4 OR EPLG4 OR LEK4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_Taxid=9606;  
 [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=95140419; PubMed=783529;  
 RA Kozlovsky C.J., Maraskovsky E., McGrew J.T., Vandenbos T.,  
 RA Teepe M., Lyman S.D., Srinivasan S., Fletcher F.A., Gayle R.B. III,  
 RA Cerretti D.P., Beckmann M.P.;  
 RT "Ligands for the receptor tyrosine kinases hek and elk: isolation of  
 RT cDNAs encoding a family of proteins."  
 RL Oncogene 10:1239-306 (1995).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RX TISSUE=B-cell;  
 RX MEDLINE=20076261; PubMed=10607706;  
 RA Aasheim H.C., Munthe E., Funderud S., Smeland E.B., Betske K.,  
 RA Logtenberg T.;  
 RT "A splice variant of human ephrin-A4 encodes a soluble molecule that  
 RT is secreted by activated human B lymphocytes."  
 RL Blood 95:1221-230 (2000).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE INTERACTION BETWEEN ACTIVATED B  
 CC LYMPHOCYTES AND DENDRITIC CELLS IN TONSILLS.  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR  
 CC OR SECRETED (DEPENDING ON THE ISOFORM).  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1/GPI-anchored (shown here) and  
 CC 2/secreted; are produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE ADULT SPLEEN, LYMPH NODE,  
 CC PROSTATE, OVARY, SMALL INTESTINE, AND COLON, AND IN FETAL HEART,  
 CC LUNG, LIVER AND KIDNEY. ALSO DETECTED IN HEMATOPOIETIC CELL LINES.  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U14188; AAC50079.1; -  
 DR EMBL; AJ006352; CAA06992.1; -  
 DR EMBL; AJ006353; CAA06993.1; -  
 DR Genew; HGNC:3224; EPRNA4.  
 DR MTW; 601380; -  
 DR InterPro; IPR001799; Ephrin.  
 DR Pfam; PF00812; Ephrin; 1.  
 DR PRINTS; PRO1347; EPHRIN.  
 DR PRODOM; PD002533; Ephrin; 1.  
 DR PROSITE; PS01299; EPHRIN; 1.  
 KM Glycoprotein; GPI-anchor; Signal; Alternative splicing.  
 FT SIGNAL 1 22  
 FT CHAIN 1 170  
 FT PROPEP 171 201  
 FT CARBOHYD 33 33  
 FT LIPID 170 170  
 FT REMOVED IN NATURE FORM (POTENTIAL).  
 FT N-LINKED (GLCNAC...) (POTENTIAL).  
 FT GPI-ANCHOR (POTENTIAL).

FT VARSPLIC 157 201 KSSAHFVGSFGESCTGMRGDTPPSLCLLLLLILRL  
 FT FT LRL -> NLPFSHPKEPSSQDPLEEGSLPALGVPIQTD  
 FT SQ SEQUENCE 201 AA; 22386 MM; ABE8D5443A9AF28D CRC64;  
 Query Match 9.2%; Score 169.5; DB 1; Length 201;  
 Best Local Similarity 29.9%; Pred. No. 0.00012;  
 Matches 66; Conservative 18; Mismatches 82; Indels 55; Gaps 10;

QY 24 LVSGLSL-EPYVNSANKRFQAEGLVLPQIGRLDLCPRAPPGHSPNTEFYLY 82  
 DB 20 LRGGSSLRHHVYVWNSNRL-LRGDAVELGLINDYLDIVCHYECPGPEGP--ETFAly 76  
 QY 83 LVGAQGRCEAP-PAPVLLTCDRPDLAPRTIKFQVSYNLMHERSHHDYLLIITS 141  
 DB 77 VDWDPGYESCGAEGRAYKRWCSLPFGHVPSKIQRFPTPSLGFELPGERTYIYISVP 136  
 QY 142 DGTREGLSLQGGVCLTRGMKVLRVGSPFGCAVPRKPVSEMPERDGAHSLPEGKE 201  
 DB 137 --TPE-----SSGQCL-----RLQGVCCCKERSBSAHFY----- 164

QY 202 NLPDPTSNATS--RGAGPLPPSPMPAVAGAAGLALLL 240  
 DB 165 --GSPGESGTSGWRGDPFPSP-----LCULLL 189

RESULT 18  
 EF45\_BRARE STANDARD; PRT; 228 AA.

AC P79728;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ephrin-A5 precursor (EPH-related receptor tyrosine kinase ligand 7)  
 DE (LEK-7) (AL-1) (ZEPHRA4)  
 GN EFNA5 OR EFNA5B OR AL1 OR EPLG7 OR LEK7.  
 OS Brachydanio rerio (zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OX NCBI\_Taxid=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=embryo;  
 RX MEDLINE=97195707; PubMed=9043080;  
 RA Brennan C., Monschau B., Lindberg R., Guthrie B., Drescher U.,  
 RA Bonhoeffer F., Holder N.;  
 RT "Two Ebf receptor tyrosine kinase ligands control axon growth and may  
 RT be involved in the creation of the retinotectal map in the  
 RT zebrafish."  
 RL Development 124:655-664 (1997).  
 CC -1- FUNCTION: CONTROL AXON GROWTH AND MAY BE INVOLVED IN THE CREATION  
 CC OF THE RETINO-TECTAL MAP.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
 CC (potential).  
 CC -1- TISSUE SPECIFICITY: WIDESPREAD EXPRESSION IN THE EMBRYO.  
 CC -1- DEVELOPING EMBRYOS FROM THE SIX-SOMITE STAGE. BY 24 HOURS IT IS  
 CC STRONGLY EXPRESSED IN THE MIDBRAIN CAUDAL TO THE PRESUMPTIVE  
 CC TECTUM. AT LATER STAGES IT IS MAINTAINED AT THE POSTERIOR MARGIN  
 CC OF THE TECTUM.  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Y09669; CAA70864.1; -  
 DR ZFIN; ZDB-GENE-980526-186; efna5b.





RA Feuerstein C., Robbins S.M.;  
 RT "Compartmentalized signaling by GPI-anchored ephrin-A5 requires the  
 RL Fyn tyrosine kinase to regulate cellular adhesion."  
 CC Genes Dev. 13:3125-3135 (1999).  
 CC -1- FUNCTION: MAY FUNCTION ACTIVELY TO STIMULATE AXON PASCIFICATION.  
 CC INDUCES COMPARTMENTALIZED SIGNALING WITHIN A CAVEOLAE-LIKE  
 CC MEMBRANE MICRODOMAIN WHEN BOUND TO THE EXTRACELLULAR DOMAIN OF ITS  
 CC COGNATE RECEPTOR. THIS SIGNALING EVENT REQUIRES THE ACTIVITY OF  
 CC THE FYN TYROSINE KINASE.  
 CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA3 AND  
 CC EPHA1.  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR IT  
 CC IS COMPARTMENTALIZED IN DISCRETE CAVEOLAE-LIKE MEMBRANE  
 CC MICRODOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U26403; AAB60377.1; -.  
 CC Genew; HGNC:3225; EFNA5.  
 CC MIM; 601535; -.  
 CC InterPro: IPR001799; Ephrin.  
 CC Pfam; PF00812; Ephrin.1.  
 CC PRINTS; PR01347; EPHRIN.  
 CC PRODOM; PD002533; Ephrin.1.  
 CC PROSITE; PS01299; EPHRIN.1.  
 CC Developmental protein; Neurogenesis; Glycoprotein; GPI-anchor; signal;  
 CC Polymorphism.  
 CC KW SIGNAL. 1 20 POTENTIAL.  
 CC CHAIN 21 228 EPHRIN-A5.  
 CC CARBOHYD 37 37 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC VARIANT 55 55 N->K (IN DBSNP:469062).  
 CC FT VARLANT /FTID=VAR\_012035.  
 CC FT SEQUENCE 228 AA; 26297 MW; 6893B1CCACFF3F57 CRC64;  
 CC SQ  
 CC Query Match 9.1%; Score 167.5; DB 1; Length 228;  
 CC Best Local Similarity 28.8%; Pred. No. 0.00019;  
 CC Matches 65; Conservative 29; Mismatches 81; Indels 51; Gaps 11;  
 CC  
 CC QY 33 VYVNSANKRFQAEAGVYLPQIGRLDILCPARPSPGHSSPNVEFYKLYLVG----- 85  
 CC DB 34 VYVNSNSNPRFQ-RGDYHIDVCINDYLDVFCPHYEDSVPEDKT--ERYVLVYVNFDSYAC 90  
 CC QY 86 -----GAQGRRCAPAPNLLTCRDLDTFTTKFOYSFNLMGHFRSHDYIYIAT 140  
 CC DB 91 DHTSKGFKMECNRPSPN-----GPIKFSKFLQFTFSLGFRRGREFYIIS 141  
 CC QY 141 S---DGTREGLESLOGVCLTRGMKYLRLVQSPRGAAVPRKPVSEMERDRGAHSL 197  
 CC DB 142 AIPDNGRRS-----CLK--LKVFVR-----PTNSCKMTIGVDRVVDNDKVENSL 186  
 CC QY 198 PKENLPDPTSNATSGAEGPLPPSPMPAVAGAAGLALLILGVA 243  
 CC DB 187 PADDTV---HESAEPSRG-ENNAQTPTIPSRL-----LAILFLLA 223  
 CC  
 CC RESULT 21  
 CC EFNA5\_MOUSE STANDARD; PRT; 228 AA.  
 CC ID EFNA5\_MOUSE  
 CC AC 008543; 008544;  
 CC DT 01-NOV-1997 (Rel. 35; Created)  
 CC DT 01-NOV-1997 (Rel. 35; Last sequence update)  
 CC DT 30-MAY-2000 (Rel. 39; Last annotation update)  
 CC DE Ephrin-A5 precursor (Eph-related receptor tyrosine kinase ligand 7)  
 CC (LEAK-7) (AL-1).  
 CC GN EFNA5 OR EPLG7 OR LEK7 OR EPL7.  
 CC OS Mus musculus (Mouse).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muride; Murinae; Mus.  
 CC NCBI\_taxid=10090;  
 CC RN (1)  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE=97060319; PubMed=8903354;  
 CC RA Flemken A.M., Gale N.W., Yancopoulos G.D., Wilkinson D.G.;  
 CC "Distinct and overlapping expression patterns of ligands for  
 CC Eph-related receptor tyrosine kinases during mouse embryogenesis";  
 CC Dev. Biol. 179:382-401 (1996).  
 CC RL Dev. Biol. 179:382-401 (1996).  
 CC CC -1- FUNCTION: MAY FUNCTION ACTIVELY TO STIMULATE AXON PASCIFICATION.  
 CC INDUCES COMPARTMENTALIZED SIGNALING WITHIN A CAVEOLAE-LIKE  
 CC MEMBRANE MICRODOMAIN WHEN BOUND TO THE EXTRACELLULAR DOMAIN OF ITS  
 CC COGNATE RECEPTOR. THIS SIGNALING EVENT REQUIRES THE ACTIVITY OF  
 CC THE FYN TYROSINE KINASE (BY SIMILARITY).  
 CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA3 AND  
 CC EPHA1 (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR IT  
 CC IS COMPARTMENTALIZED IN DISCRETE CAVEOLAE-LIKE MEMBRANE  
 CC MICRODOMAINS (BY SIMILARITY).  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U90664; AAB50239.1; -.  
 CC EMBL; U90665; AAB50240.1; -.  
 CC MGI; MGI:107444; Efna5.  
 CC DR InterPro: IPR001799; Ephrin.  
 CC Pfam; PF00812; Ephrin.1.  
 CC PRINTS; PR01347; EPHRIN.  
 CC PRODOM; PD002533; Ephrin.1.  
 CC PROSITE; PS01299; EPHRIN.1.  
 CC Developmental protein; Neurogenesis; Glycoprotein; GPI-anchor; signal;  
 CC Alternative splicing.  
 CC KW ALTERNATIVE SPLICING.  
 CC CHAIN 1 20 POTENTIAL.  
 CC FT CHAIN 21 228 EPHRIN-A5.  
 CC FT CARBOHYD 37 37 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC FT VARSPLIC 163 189 MISSING (IN SHORT ISOFORM).  
 CC SQ SEQUENCE 228 AA; 26339 MW; 85439F5337420022 CRC64;  
 CC  
 CC Query Match 9.1%; Score 167.5; DB 1; Length 228;  
 CC Best Local Similarity 28.8%; Pred. No. 0.00019;  
 CC Matches 65; Conservative 29; Mismatches 81; Indels 51; Gaps 11;  
 CC  
 CC QY 33 VYVNSANKRFQAEAGVYLPQIGRLDILCPARPSPGHSSPNVEFYKLYLVG----- 85  
 CC DB 34 VYVNSNSNPRFQ-RGDYHIDVCINDYLDVFCPHYEDSVPEDKT--ERYVLVYVNFDSYAC 90  
 CC QY 86 -----GAQGRRCAPAPNLLTCRDLDTFTTKFOYSFNLMGHFRSHDYIYIAT 140  
 CC DB 91 DHTSKGFKMECNRPSPN-----GPIKFSKFLQFTFSLGFRRGREFYIIS 141  
 CC QY 141 S---DGTREGLESLOGVCLTRGMKYLRLVQSPRGAAVPRKPVSEMERDRGAHSL 197  
 CC DB 142 AIPDNGRRS-----CLK--LKVFVR-----PTNSCKMTIGVDRVVDNDKVENSL 186  
 CC QY 198 PKENLPDPTSNATSGAEGPLPPSPMPAVAGAAGLALLILGVA 243  
 CC DB 187 PADDTV---HESAEPSRG-ENNAQTPTIPSRL-----LAILFLLA 223  
 CC  
 CC RESULT 22  
 CC EFNA5\_RAT STANDARD; PRT; 228 AA.  
 CC ID EFNA5\_RAT  
 CC AC P97605;

DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Ephrin-A5 precursor (EPH-related receptor tyrosine kinase ligand 7)  
 DE (Lerk-7) (AL-1)  
 GN EPN5 OR EPL67 OR LERK7.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=Sprague-Dawley;  
 RX MEDLINE=95267434; PubMed=7748564;  
 RA Winslow J.W., Moran P., Valverde J., Shih A., Yuan J.Q., Wong S.C.,  
 RA Tsai S.P., Goddard A., Henzel W.J., Hefti F., Beck K.D., Caras I.W.;  
 RT "Cloning of AL-1, a ligand for an Eph-related tyrosine kinase  
 RT receptor involved in axon bundle formation.";  
 RL Neuron 14:973-981(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RA Li Y.Y., McTierman C.F., Feldman A.M.;  
 RT LERK7, rat ligand for Eph-related receptor tyrosine kinase.";  
 RL Submitted (Mar-1998) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: MAY FUNCTION ACTIVELY TO STIMULATE AXON FASCICULATION.  
 CC INDICES COMPARTMENTALIZED SIGNALING WITHIN A CAVEOLAR-LIKE  
 CC MEMBRANE MICRODOMAIN WHEN BOUND TO THE EXTRACELLULAR DOMAIN OF ITS  
 CC COGNATE RECEPTOR. THIS SIGNALING EVENT REQUIRES THE ACTIVITY OF  
 CC THE FYN TYROSINE KINASE (BY SIMILARITY).  
 CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA3 AND  
 CC EPHB1 (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR IT  
 CC IS COMPARTMENTALIZED IN DISCRETE CAVEOLAE-LIKE MEMBRANE  
 CC MICRODOMAINS (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, HEART, PLACENTA AND LUNG.  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U69279; AAC05801.1; -  
 DR InterPro: IPR001799; Ephrin.  
 DR Pfam: PF00812; Ephrin; 1.  
 DR PRINTS: PR01347; EPHRIN.  
 DR Prodom: PD002533; Ephrin; 1.  
 DR PROSITE: PS01299; EPHRIN; 1.  
 KW Developmental protein; Neurogenesis; Glycoprotein; GPI-anchor; Signal.  
 FT SIGNAL 1 20  
 FT CHAIN 1 228  
 FT CARBOHYD 37 37 N-LINKED (GLCNAC... ) (POTENTIAL).  
 SQ SEQUENCE 228 AA; 26358 MW; 85598532D580022 CRC64;  
 Query Match 9.1%; Score 167.5; DB 1; Length 228;  
 Best Local Similarity 28.8%; Pred. No. 0.00019;  
 Matches 65; Conservative 29; Mismatches 81; Indels 51; Gaps 11;  
 QY 33 VYVNSANKRFOAEGGYVLYPQIGRLDLLCPARAPPGPHSSPNYEFYKLYLVG----- 85  
 DB 34 VYVNSNSNPRQ-RGDYHIDVICNDYDVCPHYEDSVPEDKT--EYVLYVNFQGYSNAC 90  
 QY 86 -----GAQGRRCAPAPNLLFTCDRPPDLKRTTFQGYSPVLMGHERSHDYIAT 140  
 DB 91 DHTSKGFKEKRCRPHSPN-----GPKKSEKFOLETFPSIGFEFRPREPYFS 141  
 QY 141 S---DGTREGLEBLGGVCLTRGKATLVRVGGSPRGCAVPKRPVSEMPERDRGAASILE 197  
 DB 142 AITDNGRRS-----CLK--LKYFVR-----PTMSCKKITGVADRVADVNDKXENBLE 186

QY 198 FGKENVLPDPTSNATSKAGGPIPPSPMPAVAGAGLALLLGVA 243  
 DB 187 PADDTV---HESAPSPRG-ENNAQTPIRPERL-----LAILLFLLA 223  
 RESULT 23  
 EFAL\_HUMAN STANDARD; PRT; 205 AA.  
 ID EFAL\_HUMAN  
 AC P20827;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DE Ephrin-A1 precursor (EPH-related receptor tyrosine kinase ligand 1)  
 DE (Lerk-1) (Immediate early response protein B61) (Tumor necrosis  
 DE factor, alpha-induced protein 4).  
 GN EFNA1 OR EPL1 OR LERK1 OR TNFAIP4.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91042512; PubMed=2233719;  
 RA Holzman L.B., Marks R.M., Dixie V.M.;  
 RT "A novel immediate-early response gene of endothelium is induced by  
 RT cytokines and encodes a secreted protein.";  
 RL Mol. Cell. Biol. 10:5830-5838(1990).  
 RN [2]  
 RP GPI-ANCHOR.  
 RX MEDLINE=95140419; PubMed=7838529;  
 RA Kozlowsky C.J., Maraskovsky E., McGrew J.T., Vandendos T.,  
 RA Teepe W., Lyman S.D., Srinivasan S., Fletcher P.A., Gayle R.B. III,  
 RA Carrecci D.P., Beckmann M.P.;  
 RT "Ligands for the receptor tyrosine kinases hek and elk: isolation of  
 RT cDNAs encoding a family of proteins.";  
 RL Oncogene 10:299-306(1995).  
 CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA4,  
 CC EPHA5, EPHA6 AND EPHA7. ALSO BINDS WITH LOW AFFINITY TO EPHA1.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- INDUCTION: BY TNF-ALPHA AND INTERLEUKIN-1 BETA.  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M57730; AA58388.1; -  
 DR PIR: A36377; A36377.  
 DR Genew; HGNC:3221; EFNA1.  
 DR MIM; 191164; -  
 DR InterPro: IPR001799; Ephrin.  
 DR Pfam: PF00812; Ephrin; 1.  
 DR PRINTS: PR01347; EPHRIN.  
 DR Prodom: PD002533; Ephrin; 1.  
 DR PROSITE: PS01299; EPHRIN; 1.  
 KW Glycoprotein; GPI-anchor; Signal.  
 FT SIGNAL 1 17  
 FT CHAIN 1 26  
 FT CARBOHYD 18 26 N-LINKED (GLCNAC... ) (POTENTIAL).  
 SQ SEQUENCE 205 AA; 23771 MW; 4FEBC6BF4C1251A9 CRC64;  
 Query Match 9.0%; Score 166; DB 1; Length 205;  
 Best Local Similarity 27.5%; Pred. No. 0.00021;  
 Matches 46; Conservative 31; Mismatches 74; Indels 16; Gaps 5;  
 QY 18 LLAGTGLVSLGLESPVYVNSANKRFOAEGGYVLYPQIGRLDLLCPARAPPGPHSSPN-- 75  
 DB 8 LLAGTGLVSLGLESPVYVNSANKRFOAEGGYVLYPQIGRLDLLCPARAPPGPHSSPN-- 75

QY 76 YEFKLYVGGAGRCRCEAPAPNLLTCDPDI---DLRFITKEQYSNLMGHEFRSH 132  
 DB 64 MEQYILYLVHEHEEYOLCOPSKDQVRWCNPNPAKHPEKXSEKFRFTFTLGEKFEKEG 123  
 QY 133 HDYIITSDGTREGLESLOGVCLTRGMKYLIVGSGPGANVPRK 119  
 DB 124 HSYIYSKPIYHEDR-----CLRLKVTYSGKITHSFQAHVNPDE 163

RESULT 24  
 ID\_EFAL\_MOUSE STANDARD; PRT; 205 AA.  
 AC P9753;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Ephrin-A1 precursor (EPH-related receptor tyrosine kinase ligand 1)  
 DE (LERK-1) (Immediate early response protein B61).  
 GN EFNA1 OR EPGI1 OR LERK1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Mistar; TISSUE=Brain;  
 RX MEDLINE=95405853; PubMed=7675446;  
 RA Takahashi H., Ikeda T.;  
 RT "Molecular cloning and expression of rat and mouse B61 gene:  
 RT implications on organogenesis.";  
 RL Oncogene 11:879-883(1995).  
 CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA4,  
 CC EPHAS, EPHA6 AND EPHA7. ALSO BINDS WITH LOW AFFINITY TO EPHA1 (BY  
 CC SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By  
 CC similarity).  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; D38056; BAA07242.1; -  
 DR InterPro; IPR001799; Ephrin.  
 DR Pfam; PF00812; Ephrin; 1.  
 DR PRINTS; PRO1347; EPHRIN.  
 DR ProDom; PD002533; Ephrin; 1.  
 DR PROSITE; PS01299; EPHRIN; 1.  
 KW Glycoprotein; GPI-anchor; Signal.  
 FT SIGNAL 1 17  
 FT CHAIN 18 205  
 FT CARBOHYD 26 26  
 FT SEQUENCE 205 AA; 23718 MW; CB6DAB3DB56A6EAD CRC64;  
 SQ

Query March 8.8%; Score 162.5; DB 1; Length 205;  
 Best local similarity 25.7%; Pred. No 0.00036;  
 Matches 49; Conservative 33; Mismatches 86; Indels 23; Gaps 6;

QY 18 LIGVLGVSGLSLEPVYVNSANKRFQAGGYLVYQIGDRLLICPRAPPGPHSSPNYE 77  
 DB 8 LIGLCCSLAADVDRHIVFNWSSNPKFRRE-DYTVHQLNDYLDIICPHYEDSV-ADAAVE 65  
 QY 78 FYKLYIVGAGRCRCEAPAPNLLTCDPDI---DLRFITKEQYSNLMGHEFRSH 134  
 DB 66 RYSLVWVHQEVVTEPOSQDVQVRKCNQPSAKGPEKLSKFRFTFTLGEKFEKGS 125  
 QY 135 YIATSDGTREGLESLOGVCLTRGMKYLIVGSGPGANVPRKPVSEMPMERD----- 189

DB 126 YIYKSPYHDE-----TQCLKXVTWNGKITHSFAHVPQ-----EKLQADDEPQ 174  
 QY 190 --RGAHSLSP 198  
 DB 175 VLHSGHSAAP 185

RESULT 25  
 ID\_EFAL\_MOUSE STANDARD; PRT; 205 AA.  
 AC P52793; P97331;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 35, Last annotation update)  
 DE Ephrin-A1 precursor (EPH-related receptor tyrosine kinase ligand 1)  
 DE (LERK-1) (Immediate early response protein B61).  
 GN EFNA1 OR EPGI1 OR LERK1 OR EPL1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ICR;  
 RX MEDLINE=95405853; PubMed=7675446;  
 RA Takahashi H., Ikeda T.;  
 RT "Molecular cloning and expression of rat and mouse B61 gene:  
 RT implications on organogenesis.";  
 RL Oncogene 11:879-883(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c;  
 RA Morris J.C., Chiarletta A., Morris G.E., Giannotti J., Caruso A.,  
 RA Hammett D.J., Finnerty H., Turner K., Wood C.R.;  
 RL Submitted (May-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97060319; PubMed=8903354;  
 RA Flemmiken A.M., Gale N.W., Yancopoulos G.D., Wilkinson D.G.;  
 RT "Distinct and overlapping expression patterns of ligands for  
 RT Eph-related receptor tyrosine kinases during mouse embryogenesis.";  
 RL Dev. Biol. 179:382-401(1996)  
 CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA4,  
 CC EPHAS, EPHA6 AND EPHA7. ALSO BINDS WITH LOW AFFINITY TO EPHA1.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By  
 CC similarity).  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; D38146; BAA07344.1; -  
 DR EMBL; U26188; AAA67563.1; -  
 DR EMBL; U90662; AAB50237.1; -  
 DR MGD; MGI:103236; Efnal.  
 DR InterPro; IPR001799; Ephrin.  
 DR Pfam; PF00812; Ephrin; 1.  
 DR PRINTS; PRO1347; EPHRIN.  
 DR ProDom; PD002533; Ephrin; 1.  
 DR PROSITE; PS01299; EPHRIN; 1.  
 KW Glycoprotein; GPI-anchor; Signal.  
 FT SIGNAL 1 17  
 FT CHAIN 18 205  
 FT CARBOHYD 26 26  
 FT SEQUENCE 205 AA; 23718 MW; CB6DAB3DB56A6EAD CRC64;  
 SQ

Query March 8.8%; Score 162.5; DB 1; Length 205;  
 Best local similarity 25.7%; Pred. No 0.00036;  
 Matches 49; Conservative 33; Mismatches 86; Indels 23; Gaps 6;

QY 18 LIGVLGVSGLSLEPVYVNSANKRFQAGGYLVYQIGDRLLICPRAPPGPHSSPNYE 77  
 DB 8 LIGLCCSLAADVDRHIVFNWSSNPKFRRE-DYTVHQLNDYLDIICPHYEDSV-ADAAVE 65  
 QY 78 FYKLYIVGAGRCRCEAPAPNLLTCDPDI---DLRFITKEQYSNLMGHEFRSH 134  
 DB 66 RYSLVWVHQEVVTEPOSQDVQVRKCNQPSAKGPEKLSKFRFTFTLGEKFEKGS 125  
 QY 135 YIATSDGTREGLESLOGVCLTRGMKYLIVGSGPGANVPRKPVSEMPMERD----- 189

```

FT CONFLICT 94 94 R -> Q (IN REF. 1)
FT CONFLICT 112 112 T -> S (IN REF. 1)
FT CONFLICT 115 115 I -> T (IN REF. 1)
FT CONFLICT 138 138 S -> T (IN REF. 1)
FT CONFLICT 154 154 N -> S (IN REF. 1)
FT CONFLICT 156 156 Q -> H (IN REF. 1)
FT CONFLICT 159 159 V -> A (IN REF. 1)
FT CONFLICT 181 181 Y -> H (IN REF. 1)
FT CONFLICT 204 204 S -> T (IN REF. 1)
SQ SEQUENCE 205 AA; 23802 MW; 5ABF3AEE2091E868 CRC64;

Query Match
Best Local Similarity 26.1%; Score 161; DB 1; Length 205;
Matches 43; Conservative 33; Mismatches 77; Indels 12; Gaps 4;

QY 18 ILGLVGLVSGLSLPEYVNSANKRFOAGGYVLPQIGRLDLCPRARPPGHSPPNTE 77
DB 8 LILGLCSLAAADRHIVFMNNSNPKFRB-DYIVHVLNDYLDICPHYEDDSV-ADAAHE 65
QY 78 FYULYVGAQGRCEAPPAVNLITCDRDL--DLRFTIKQEQYSPLMGHFRSHHD 134
DB 66 RYLYLVWEHQEVYACQPOSKQVRMNCNRPSPAKGPEKLSKQFRTFPLIGKFEKHS 125
QY 135 YVLIATSDGTRREGLESLQGVCLTRGKVKYLIRYQSPRGAVPRK 179
DB 126 YVYISKPIYHOE-----SQCLKLVYNGKITNHPQAHVNFQE 163

RESULT 26
EPF4_MOUSE STANDARD; PRT; 206 AA.
ID ID EPF4_MOUSE STANDARD; PRT; 206 AA.
AC 008542; 055218; (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ephrin-A4 precursor (EPH-related receptor tyrosine kinase ligand 4)
GN (LERK-4).
OS ERN4 OR EPUG4 OR LERK4 OR EPL4.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97060319; PubMed=8903354;
RA Flemken A.M., Gale N.W., Yancopoulos G.D., Wilkinson D.G.;
RT "Distinct and overlapping expression patterns of ligands for
RT Eph-related receptor tyrosine kinases during mouse embryogenesis."
RT Dev. Biol. 179:382-401 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=98126446; PubMed=9465306;
RA Cerretti D.P., Nelson N.;
RT "Characterization of the genes for mouse LERK-3/Ephrin-A3 (Epl3),
RT mouse LERK-4/Ephrin-A4 (Epl4), and human LERK-6/Ephrin-A2 (Epl6):
RT conservation of intron/exon structure."
RT Genomics 47:131-135 (1998).
RN [3]
RP SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
RN similarity).
CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U90663; AAB50238.1; -
DR EMBL; U92890; AAC39962.1; -
DR EMBL; U92889; AAC39962.1; JOINED.

```

```

DR MGI:106643; Efn4.
DR InterPro; IPR001799; Ephrin.
DR Pfam; PF00812; Ephrin.1.
DR PRINTS; PR01347; EPHRIN.
DR PRODOM; PD002533; Ephrin.1.
DR PROSITE; PS01299; EPHRIN.1.
KM Glycoprotein; GPI-anchor; Signal.
FT SIGNAL 1 25
FT CHAIN 26 206 EPHRIN-A4.
FT CARBOHYD 33 33 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SITE 41 43 CELL ATTACHMENT SITE (POTENTIAL).
FT CONFLICT 1 4 MRL -> MLRLGLYPTPRPAPPPV (IN REF. 1).
SQ SEQUENCE 206 AA; 22861 MW; 43501971DDIC6EAS CRC64;

Query Match
Best Local Similarity 28.1%; Score 160.5; DB 1; Length 206;
Matches 61; Conservative 16; Mismatches 81; Indels 59; Gaps 10;

QY 32 PYVNSANKRFOAGGYVLPQIGRLDLCPRARPPGHSPPNTEFYKLYVGAQGR 91
DB 29 PYVNSNPLR-LRGAVALVGLFNDYLDICPHYESPGPEGF--ETPALYMDVMSGYEA 85
QY 92 CEAPPAVNLITCDRDL--DLRFTIKQEQYSPLMGHFRSHHDYIATSDGTRREG 147
DB 86 CTAGANAFORWNCMPFAPSPVRFSEKIQRVTFPLGFEFLGRTYIYISVTPESP 145
QY 148 -LESLOGVCLTRGKVKYLIRYQSPRGAVPRKVSMPMRDGAH-GLERKENTLPG 205
DB 146 RCLLQVSVCC-----KEGSSHEANP-----VG 170
QY 206 DPTSNATS--RGAEGLPPSPMPAVAGAGALLL 240
DB 171 SPGESGTSGMRGAPSP-----LCILL 194

RESULT 27
CALA_MOUSE STANDARD; PRT; 680 AA.
ID ID CALA_MOUSE STANDARD; PRT; 680 AA.
AC 005306;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Collagen alpha 1(X) chain precursor.
GN COL10A1.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RX MEDLINE=93143676; PubMed=8424763;
RA Elima K., Berola I., Rosati R., Metsaranta M., Garofalo S., Perala M.,
RA de Crombrughe B., Vuorio E.;
RT "The mouse collagen X gene: complete nucleotide sequence, exon
RT structure and expression pattern."
RT Biochem. J. 289:247-253 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV; TISSUE=Liver;
RX MEDLINE=9328750; PubMed=8477738;
RA Kong R.Y.C., Kwan K.M., Lau E.T., Thomas J.T., Boot-Handford R.P.,
RA Grant M.E., Cheah K.S.E.;
RT "Intron-exon structure, alternative use of promoter and expression of
RT the mouse collagen X gene, Col10a-1."
RT Eur. J. Biochem. 213:99-111 (1993).
RN [3]
RP SEQUENCE OF 51-680 FROM N.A.
RC STRAIN=DRA/2J;
RX MEDLINE=92267014; PubMed=1587271;
RA Apfe S.S., Seidlin M.F., Hayaishi M., Olsen B.R.;

```

RT "Cloning of the human and mouse type X collagen genes and mapping of  
 the mouse type X collagen gene to chromosome 10.";  
 RL Eur. J. Biochem. 206:217-224(1992).  
 RP SEQUENCE OF 385-627 FROM N.A.  
 RC STRAIN=CS7BL/6;  
 RX MEDLINE=92182017; PubMed=1543751;  
 RA Lima K., Metsaeranta M., Kallio J., Peraelae M., Seroia I.,  
 RA Garofalo S., de Crombrughe B., Vuorio E.;  
 RT "Specific hybridization probes for mouse alpha 2(I)X and alpha 1(X)  
 collagen mRNAs.";  
 RL Biochim. Biophys. Acta 1130:78-80(1992).  
 CC -1- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTROPHIC  
 CC CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE  
 CC MINERALIZATION ZONES OF HYALINE CARTILAGE.  
 CC -1- SUBUNIT: HOMOTRIMER.  
 CC -1- PMW: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC -1- UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: STRONG, TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.  
 CC -1- SIMILARITY: CONTAINS 1 C1Q DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X67348; CAA47763.1; -;  
 DR EMBL; X65121; CAA46237.1; -;  
 DR EMBL; X63013; CAA44741.1; -;  
 DR EMBL; Z21610; CAA79736.1; -;  
 DR PIR; S28807; S28807.  
 DR PIR; S31216; S31216.  
 DR PIR; S22215; S22215.  
 DR MED; MGI:88445; C011041.  
 DR InterPro; IPR001073; C1Q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1q; 1.  
 DR Pfam; PF01391; Collagen; 9.  
 DR PRINTS; PRO0007; COMPLEMENTC1Q.  
 DR PRODOM; PD000007; Collagen; 2.  
 DR SMART; SMO0110; C1Q; 1.  
 DR PROSITE; PS01113; C1Q; 1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Cartilage; Collagen; Signal.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT CHAIN 1 680 COLLAGEN ALPHA 1(X) CHAIN.  
 FT DOMAIN 19 56 NONHELICAL REGION (NC2).  
 FT DOMAIN 57 519 TRIPLE-HELICAL REGION.  
 FT DOMAIN 520 680 NONHELICAL REGION (NC1).  
 FT DOMAIN 545 680 C1Q.  
 FT DOMAIN 248 248 P -> L (IN REF. 3).  
 FT CONFLICT 286 286 A -> S (IN REF. 2).  
 FT CONFLICT 306 306 L -> F (IN REF. 3).  
 FT CONFLICT 417 417 T -> S (IN REF. 3).  
 FT CONFLICT 451 451 R -> K (IN REF. 4).  
 FT CONFLICT 500 500 H -> L (IN REF. 3).  
 FT CONFLICT 567 567 A -> C (IN REF. 3).  
 FT CONFLICT 569 569 I -> H (IN REF. 3).  
 FT CONFLICT 572 572 FD -> Y (IN REF. 3).  
 FT CONFLICT 635 635 Q -> T (IN REF. 3).  
 SQ SEQUENCE 680 AA; 66775 MW; PE984CA99AF08B2 CRC64;  
 Query Match 8.6%; Score 159; DB 1; Length 680;  
 Best Local Similarity 25.5%; Pred. No. 0.002;  
 Matches 97; Conservative 30; Mismatches 108; Indels 146; Gaps 24;  
 QY 1 MGPPSGGPGGV-RVGAALLLGLVGLVSGLSLEPYVANSANKRPAQAGGYLVLPQIGRLD 59  
 Db 211 IGPP--GPGSGRGGRGNGPFGQPGI-----KGRNGFGENG----- 244

QY 60 ILCPARPSPSPSPNVEFYKLVVGAQGRCEAPAPNLLLTCDRDPDLRTIKFOE 119  
 Db 245 ----FSGPPGPGPP-----GKGR--EIGKPGALISGCGGI----- 277  
 QY 120 YSPNIMGEHFRSHDHYIATSDGT---REGLESLOGVCLTRMKYLLRVGSGPRGGA 175  
 Db 278 --PGEKGPAGP-----IAGPPGAPGPKQGLPGLRG---QRG-----PAG-- 314  
 QY 176 VPKRVSSEMPKRDGAHSLSPGKENTPGDPTSNATRGAGGLPPSPMDVAGAGSL 235  
 Db 315 LPGAFA---KGERGPPG--HPGPGLPSP---GNMGPGPKGIPGHHGIPAKKEI 364  
 QY 236 ALLLVAGAGAGMCMRRRAKP---SESRHPG-----PSGFRGSLGVGGGG 282  
 Db 365 G--LVGAPGPFA---RGARPGGLDGKTYGPEBGLNGPKXNPLPQKQDPPGVGTGPG 419  
 QY 283 M-----GPRAREPGLGIALRGGAADPPPHXKXSGDVGHVY 323  
 Db 420 LRGPPVPGAKGVPQGNBAGPR-GEPIPGTR---GTPGPPGVGPPSGKSDPGNF-- 472  
 QY 324 IVQDGP-----POSPP 334  
 Db 473 -GAPPGATATKGLNGPFGPP 492  
 RESULT 28  
 ID CAL3 BOVIN STANDARD; PRT; 1049 AA.  
 AC P04258;  
 DT 20-MAR-1987 (Rel. 04, Created)  
 DT 20-MAR-1987 (Rel. 04, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 1(III) chain.  
 GN COL3A1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 ON NCBI\_TaxID=9913;  
 RX MEDLINE=80026026; PubMed=488906;  
 RA Fietzek P.P., Allmann H., Rautenberg J., Henkel W., Wächter E.,  
 RA Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. I. The amino  
 RT acid sequence of the amino terminal region of the alpha 1(III) chain  
 RT (positions 1-222)."  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:809-820(1979).  
 RN [2]  
 RP SEQUENCE OF 243-422.  
 RX MEDLINE=80026027; PubMed=488907;  
 RA Dewes H., Fietzek P.P., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. II. The amino  
 RT acid sequence of the cyanogen bromide peptide alpha 1(III)CB1,8,10,2  
 RT (positions 223-402)."  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:821-832(1979).  
 RN [3]  
 RP SEQUENCE OF 423-571.  
 RX MEDLINE=80026028; PubMed=488908;  
 RA Bentz H., Fietzek P.P., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. III. The  
 RT amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4  
 RT (positions 403-551)."  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:833-840(1979).  
 RN [4]  
 RP SEQUENCE OF 572-808.  
 RX MEDLINE=80026029; PubMed=488909;  
 RA Lang H., Garaville R.W., Fietzek P.P., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. IV. The amino  
 RT acid sequence of the cyanogen bromide peptide alpha 1(III)CB5  
 RT (positions 552-788)."  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:841-850(1979).  
 RN [5]

RP SEQUENCE OF 803-947.  
 RX MEDLINE=60026030; PubMed=488910;  
 RA Dewes H., Fietzek P.P., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. V. The amino  
 RT acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A  
 RT (position 789-927).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:851-860(1979).  
 RN [6]  
 RP SEQUENCE OF 948-1049.  
 RX MEDLINE=80026031; PubMed=488911;  
 RA Allmann H., Fietzek P.P., Glanville R.W., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. VI. The amino  
 RT acid sequence of the carboxyterminal cyanogen bromide peptide alpha  
 RT 1(III)CB9B (positions 928-1028).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:861-868(1979).  
 RN [6]  
 RP -1- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES  
 CC ALONG WITH TYPE I COLLAGEN.  
 CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE  
 CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE  
 CC ALSO CROSS-LINKED VIA HYDROXYLINES.  
 CC -1- PPM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 DR PIR; A02862; CGB07S.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF01391; Collagen\_17.  
 DR ProDom; PD000007; Collagen\_1.  
 DR PROSITE; PS01208; WFC; PARTIAL.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Collagen.  
 FT DOMAIN 1 14 NONHELICAL REGION (N-TERMINAL).  
 FT DOMAIN 15 1040 TRIPLE-HELICAL REGION.  
 FT MOD\_RES 1041 1049 NONHELICAL REGION (C-TERMINAL).  
 FT MOD\_RES 95 95 HYDROXYLATION.  
 FT MOD\_RES 107 107 HYDROXYLATION.  
 FT MOD\_RES 119 119 HYDROXYLATION.  
 FT MOD\_RES 938 938 HYDROXYLATION.  
 FT MOD\_RES 950 950 HYDROXYLATION.  
 FT CARBOHYD 107 107 O-LINKED (GAL. . .).  
 FT CARBOHYD 950 950 O-LINKED (GAL. . .).  
 FT DISULFID 1040 1040 INTERCHAIN.  
 FT DISULFID 1041 1041 INTERCHAIN.  
 SQ SEQUENCE 1049 AA; 93651 MW; 8EEC3D1066EC9A3 CRC64;  
 Query Match 8.4%; Score 154.5; DB 1; Length 1049;  
 Best Local Similarity 26.5%; Pred. No. 0.006; 74; Indels 83; Gaps 8;  
 Matches 60; Conservative 9; Mismatches 74; Indels 83; Gaps 8;  
 QY 171 PRGAVPRKVESEMPERDRGA-----AASLEFGKEN 202  
 DB 688 PAGESPAGSPGQGVKRGSGPGAGFPGGGPPGPPSGNNGPPSSGAPKDG 747  
 QY 203 LRPDPISNAT-----SRGAEPLPPPSNPVAGAAAGLALLIGVAGA 245  
 DB 748 PPGPPSGNAPGPGISGPKGDSGPPGERGAPGPPGAPGLGAG-----LTGARL 802  
 QY 246 GGMWRRRRARAK-----SESRRPG-----GSFGRGSGYLGGGGGMPREAPGE 292  
 DB 803 AGPPGMPGARGSPGQIKENKRGPSGNGRGPFGGLPLAGTAGRGRDNGPS 862  
 QY 293 LGIALRG-----GAADPPCPHYEKVSGDYGPVYIVODGPP 330  
 DB 863 DGPGRDGAAPGAKGRGNGSPGAP-----GAPGHP-----GPP 896  
 RESULT 29  
 CC02\_CABEL STANDARD; RT; 301 AA.  
 ID CC02\_CABEL  
 AC P17656;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cuticle collagen 2 precursor.

GN COL-2 OR W01B6.7.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 CX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=83050944; PubMed=7139711;  
 RA Krimer J.M., Cox G.N., Hirsch D.;  
 RT "Comparisons of the complete sequences of two collagen genes from  
 RT Caenorhabditis elegans.";  
 RL Cell 30:559-606(1982).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=85105075; PubMed=2578467;  
 RA Krimer J.M., Cox G.N., Hirsch D.;  
 RT "Expression of the Caenorhabditis elegans collagen genes col-1 and  
 RT col-2 is developmentally regulated.";  
 RL J. Biol. Chem. 260:1945-1951(1985).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Coles L.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
 CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
 CC BARRIER TO PROTECT THE NEM FROM ITS ENVIRONMENT.  
 CC -1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE  
 CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-  
 CC LINKS.  
 CC -1- SIMILARITY: BELONGS TO THE CUTICULAR COLLAGEN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; J01048; AAA27990.1; -  
 DR EMBL; V00148; CAA23464.1; -  
 DR EMBL; Z68301; CAA92620.1; -  
 DR PIR; B31219; B31219  
 DR WormPep; W01B6.7; CE03759.  
 DR InterPro; IPR002486; Col\_cuticle\_N.  
 DR Pfam; PF01391; Collagen\_2.  
 DR Pfam; PF01484; Col\_cuticle\_N\_1.  
 KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen;  
 FT SIGNAL 1 37 POTENTIAL.  
 FT CHAIN 38 301 CUTICLE COLLAGEN 2.  
 FT DOMAIN 85 97 GLY-RICH.  
 FT DOMAIN 105 134 TRIPLE-HELICAL REGION.  
 FT DOMAIN 153 176 TRIPLE-HELICAL REGION.  
 FT DOMAIN 183 212 TRIPLE-HELICAL REGION.  
 FT DOMAIN 215 282 TRIPLE-HELICAL REGION.  
 SQ SEQUENCE 301 AA; 28025 MW; 33317E3BDAC302F9 CRC64;  
 Query Match 8.2%; Score 151.5; DB 1; Length 301;  
 Best Local Similarity 28.7%; Pred. No. 0.0027;  
 Matches 50; Conservative 15; Mismatches 72; Indels 37; Gaps 7;  
 QY 168 GSPRGAVPRKPYSEMPERDRG-----AASLEFGKENLPQPT--SNATSGAE 217  
 DB 129 GNPBGASAPCEPTQPCPCPGPPGAPGPPGPPDNGPSGAPSGPPGPP 188  
 QY 218 GPLPPSPNPVAGAAAGLALLIGVAGA-----GGAMCWRRRRARAKPSRRPGP--GSFGRG 273  
 DB 189 GPAPGAGNDGAPGAPG-----PQFPASGSGGP-----GSPGAPGPPGAPGAGNDG 236

CY 274 SLGCGGGMGPR-----EAEPEELIALRGGANPPFCPEYKVG 316  
 ID CA34 HUMAN STANDARD; PRT; 1670 AA.  
 AC 001955; 09B072;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 15-JUN-1999 (Rel. 38, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 3 (IV) chain precursor (Goodpasture antigen).  
 OS COL4A3.  
 GN Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CX NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=94364994; PubMed=8083201;  
 RA Maruyama M., Leinonen A., Mochizuki T., Trygvasen K., Reeders S.T.;  
 RT "Complete primary structure of the human alpha 3(IV) collagen chain.  
 RT Coexpression of the alpha 3 (IV) and alpha 4 (IV) collagen chains in  
 RT human tissues.";  
 RL J. Biol. Chem. 269:23013-23017 (1994).  
 RN [2]  
 RP REVISIONS.  
 RA Leinonen A.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A., VARIANTS AS E-297; R-407; R-640; R-1167; E-1207;  
 RP Q-1215; S-1277; T-1330; E-1334; E-1347 AND C-1661, AND VARIANTS R-43;  
 RP E-162; Y-326; H-408; R-451; L-574; E-1269 AND P-1474.  
 RX MEDLINE=21064696; PubMed=11134255;  
 RA Heider L., Aronow C., Forester L., Cohen-Solal L., Mollet G.,  
 RA "Structure of the human type IV collagen gene COL4A3 and mutations in  
 RT autosomal Alport syndrome.";  
 RL J. Am. Soc. Nephrol. 12:97-106 (2001).  
 RN [4]  
 RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=93015826; PubMed=1400291.  
 RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;  
 RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the  
 RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially  
 RT antigenic region at the triple helix/NC1 domain junction.";  
 RL J. Biol. Chem. 267:19780-19784 (1992).  
 RN [5]  
 RP SEQUENCE OF 1453-1670 FROM N.A.  
 RX MEDLINE=91353701; PubMed=1882840;  
 RA Morrison K.E., Maruyama M., Yang-Feng T.L., Reeders S.T.;  
 RT "Sequence and localization of a partial cDNA encoding the human alpha  
 RT 3 chain of type IV collagen.";  
 RL Am. J. Hum. Genet. 49:545-554 (1991).  
 RN [6]  
 RP SEQUENCE OF 1331-1670 FROM N.A.  
 RX TISSUE=Kidney;  
 RX MEDLINE=92147878; PubMed=1737849;  
 RA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,  
 RA Pusey C.D.;  
 RT "Molecular cloning of the human Goodpasture antigen demonstrates it  
 RT to be the alpha 3 chain of type IV collagen.";  
 RL J. Clin. Invest. 89:592-601 (1992).  
 RN [7]  
 RP SEQUENCE OF 1644-1670 FROM N.A.  
 RC TISSUE=Kidney;  
 RA Ding J.;  
 RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.

RC TISSUE=Kidney;  
 RX MEDLINE=94124597; PubMed=8294492;  
 RA Feng L., Xia Y., Wilson C.B.;  
 RT "Alternative splicing of the NC1 domain of the human alpha 3 (IV)  
 RT collagen gene. Differential expression of mRNA transcripts that  
 RT predict three protein variants with distinct carboxyl regions.";  
 RL J. Biol. Chem. 269:2342-2348 (1994).  
 RN [9]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=98196854; PubMed=9537506;  
 RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioke H.,  
 RA Nimomiya Y.;  
 RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3 (IV) and  
 RT alpha4 (IV) collagen chains are arranged head-to-head on chromosome  
 RT 2q36.";  
 RL FEBS Lett. 424:11-16 (1998).  
 RN [10]  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE=93280184; PubMed=8505332;  
 RA Bernal D., Quinones S., Saus J.;  
 RT "The human mRNA encoding the Goodpasture antigen is alternatively  
 RT spliced.";  
 RL J. Biol. Chem. 268:12090-12094 (1993).  
 RN [11]  
 RP VARIANT PRO-1474.  
 RX MEDLINE=95078827; PubMed=7987301;  
 RA Lemmink H.H., Mochizuki T., van den Heuvel L.P.M.J., Schroeder C.H.,  
 RA Barrientos A., Monens L.A.H., van Oost B.A., Brunner H.G.,  
 RA Reeders S.T., Smeets H.J.M.;  
 RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal  
 RT recessive Alport syndrome.";  
 RL Hum. Mol. Genet. 3:1269-1273 (1994).  
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF  
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A "CHICKEN-WIRE"  
 CC MESHWORK TOGETHER WITH LAMININS, PROTEGLYCANS AND ENACTIN/  
 CC NIDAGEN.  
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1 (IV) -  
 CC ALPHA 6 (IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE  
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.  
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS, 1 (SHOWN HERE), 2/V AND  
 CC 3/L5; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER IN THEIR  
 CC C-TERMINAL NC1 DOMAINS.  
 CC -1- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE  
 CC COLLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,  
 CC COCHLEA, LUNG AND BRAIN.  
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS  
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE  
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY  
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL  
 CC TRIPLE-HELICAL 7S DOMAIN.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- PTM: THE ALTERNATIVE SPLICED FORM V CONTAINS AN ADDITIONAL  
 CC N-LINKED GLYCOSYLATION SITE.  
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH  
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF  
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE  
 CC IV COLLAGENS.  
 CC -1- PTM: Phosphorylated by the Goodpasture antigen-binding protein.  
 CC -1- DISEASE: ANTIBODIES AGAINST THE NC1 DOMAIN OF ALPHA3(IV) MEDIATE  
 CC THE HUMAN AUTOIMMUNE DISEASE, GOODPASTURE SYNDROME, WHICH IS  
 CC CHARACTERIZED BY HEMATURIA AND PULMONARY HEMORRHAGE.  
 CC -1- DISEASE: DEFECTS IN COL4A3 ARE ASSOCIATED WITH THE TYPE I  
 CC AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME, AN HEREDITARY  
 CC GLOMERULONEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE,  
 CC HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN  
 CC MALES AND FEMALES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
DR EMBL; X60031; CAA56335.1; -  
DR EMBL; AJ288487; CAC36101.1; JOINED.  
DR EMBL; AJ288488; CAC36101.1; JOINED.  
DR EMBL; AJ288489; CAC36101.1; JOINED.  
DR EMBL; AJ288490; CAC36101.1; JOINED.  
DR EMBL; AJ288491; CAC36101.1; JOINED.  
DR EMBL; AJ288492; CAC36101.1; JOINED.  
DR EMBL; AJ288493; CAC36101.1; JOINED.  
DR EMBL; AJ288494; CAC36101.1; JOINED.  
DR EMBL; AJ288495; CAC36101.1; JOINED.  
DR EMBL; AJ288496; CAC36101.1; JOINED.  
DR EMBL; AJ288497; CAC36101.1; JOINED.  
DR EMBL; AJ288498; CAC36101.1; JOINED.  
DR EMBL; AJ288499; CAC36101.1; JOINED.  
DR EMBL; AJ288500; CAC36101.1; JOINED.  
DR EMBL; AJ288501; CAC36101.1; JOINED.  
DR EMBL; AJ288502; CAC36101.1; JOINED.  
DR EMBL; AJ288503; CAC36101.1; JOINED.  
DR EMBL; AJ288504; CAC36101.1; JOINED.  
DR EMBL; AJ288505; CAC36101.1; JOINED.  
DR EMBL; AJ288506; CAC36101.1; JOINED.  
DR EMBL; AJ288507; CAC36101.1; JOINED.  
DR EMBL; AJ288508; CAC36101.1; JOINED.  
DR EMBL; AJ288509; CAC36101.1; JOINED.  
DR EMBL; AJ288510; CAC36101.1; JOINED.  
DR EMBL; AJ288511; CAC36101.1; JOINED.  
DR EMBL; AJ288512; CAC36101.1; JOINED.  
DR EMBL; AJ288513; CAC36101.1; JOINED.  
DR EMBL; AJ288514; CAC36101.1; JOINED.  
DR EMBL; AJ288515; CAC36101.1; JOINED.  
DR EMBL; AJ288516; CAC36101.1; JOINED.  
DR EMBL; AJ288517; CAC36101.1; JOINED.  
DR EMBL; AJ288518; CAC36101.1; JOINED.  
DR EMBL; AJ288519; CAC36101.1; JOINED.  
DR EMBL; AJ288520; CAC36101.1; JOINED.  
DR EMBL; AJ288521; CAC36101.1; JOINED.  
DR EMBL; AJ288522; CAC36101.1; JOINED.  
DR EMBL; AJ288523; CAC36101.1; JOINED.  
DR EMBL; AJ288524; CAC36101.1; JOINED.  
DR EMBL; AJ288525; CAC36101.1; JOINED.  
DR EMBL; AJ288526; CAC36101.1; JOINED.  
DR EMBL; AJ288527; CAC36101.1; JOINED.  
DR EMBL; AJ288528; CAC36101.1; JOINED.  
DR EMBL; AJ288529; CAC36101.1; JOINED.  
DR EMBL; AJ288530; CAC36101.1; JOINED.  
DR EMBL; AJ288531; CAC36101.1; JOINED.  
DR EMBL; AJ288532; CAC36101.1; JOINED.  
DR EMBL; AJ288533; CAC36101.1; JOINED.  
DR EMBL; AJ288534; CAC36101.1; JOINED.  
DR EMBL; AJ288535; CAC36101.1; JOINED.  
DR EMBL; AJ288536; CAC36101.1; JOINED.  
DR EMBL; AJ288537; CAC36101.1; JOINED.  
DR EMBL; AJ288538; CAC36101.1; JOINED.  
DR EMBL; M92993; AAA21610.1; -  
DR EMBL; S55790; AAB19637.1; -  
DR EMBL; M81379; AAA51556.1; -  
DR EMBL; L08650; AAA52044.1; -  
DR EMBL; U02519; AAA18942.1; -

Query Match 8.0%; Score 148.5; DB 1; Length 1670;  
Best Local Similarity 27.1%; Pred. No. 0.023;

Matches 105; Conservative 22; Mismatches 146; Indels 115; Gaps 23;

QY 1 MGPFSH-GPGGV-----RVGALLGLVGLVSGLSLEPVYWNANRKPQAEQGVLYLPQI 54

DB 1020 MGNFGSKGKGTGLGFPGRAGPGGLPGTHGL-QGDKEPGV-----SEG----- 1061

QY 55 GDRLLDLCPRARPPGPHSSPNEYFYKLYLVGAQGRRC-----APPA 97

DB 1062 -----TRGPPGPTGDPG-----LPGDMGXKXGMPGPPGHLGPAEGAGAGS 1105  
QY 98 PNL-LILCDRPLDIDRPT-IKFQEVYPNLMGHE-----FRSHHYIYIATSDGREGLES 150  
DB 1106 PGPSPGPKPGPHDPLGKGLIKGLIGPPGIRGPPGLPDPGPGSPGPMGRDQG-RDGLPG 1164  
QY 151 LOGGVCLTRGKQVLLRVQSPRG--GAVPRKPVSEMMERDRGAHSLPEKENLPQDPT 208  
DB 1165 PAG-----EKGETGLRAPRPGRNPGAQAK-----GDRGA-----PGPGLPG-- 1204  
QY 209 SNATSRGAEGLPPSPMAYVAGAGLALLLVAGAGACMCRRRAKPSBSHPGP-- 266  
DB 1205 -RKGMAGAGRPGPTGIRGPPGPPGLPGALIPQGTGRGP--DGSRGSPAPRPPGPG 1260  
QY 267 -----GSPRGSGSLG-LGGGCGMP--RAEPGELGIALRGGAADPPPCPHYEK 313  
DB 1261 SHVIGIKGDKSMHGPKPGPGTAGDMGPRGRIAGATGTPLP--GPRGPGP-QGFPFG 1316  
QY 314 VSGDYGHFVYIVQDP-----PQSPENI 336  
DB 1317 VKGEKGNPGLGSLGPPGPIPKGPPGV 1344

# RESULT 31

ID CAH\_MOUSE STANDARD; PRT; 1527 AA.

AC P39061; Q62002; Q61437;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DS Collagen alpha 1(XVIII) chain precursor [contains: Endostatin].  
GN COL18A1.

OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;

RN [1]

RP SEQUENCE FROM N.A. (SHORT ISOFORM).  
RC STRAIN=BALB/c; TISSUE=Liver;  
RX MEDLINE=94245707; PubMed=8186673;  
RA Rehn M.V., Hintikka E., Pihlajaniemi T.;  
RT "Primary structure of the alpha 1 chain of mouse type XVIII collagen,  
RT partial structure of the corresponding gene, and comparison of the  
RT alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen  
RT chain.";  
RT J. Biol. Chem. 269:13929-13935(1994).

RN [2]

RP SEQUENCE FROM N.A. (SHORT AND LONG ISOFORMS).  
RA Rehn M., Hintikka E., Pihlajaniemi T.;  
RT "Characterization of the mouse gene for the alpha-1 chain of type  
RT XVIII collagen (COL18A1) reveals that the three variant N-terminal  
RT polypeptide forms are transcribed from two widely separated  
RT promoters.";  
RT Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 213-1140 FROM N.A. (SHORT ISOFORM).  
RX MEDLINE=94240112; PubMed=8183894;  
RA Rehn M.V., Pihlajaniemi T.;  
RT "Alpha 1(XVIII), a collagen chain with frequent interruptions in the  
RT collagenous sequence, a distinct tissue distribution, and homology  
RT with type XV collagen.";  
RT Proc. Natl. Acad. Sci. U.S.A. 91:4234-4236(1994).

RN [4]

RP SEQUENCE OF 240-1527 FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=94240111; PubMed=8183893;  
RA Oh S.P., Kamagata Y., Muzgaki Y., Timmons S., Ooshima A., Olsen B.R.;  
RT "Isolation and sequencing of cDNAs for proteins with multiple domains  
RT of Gly-Xaa-Yaa repeats identify a distinct family of collagenous  
RT proteins.";  
RT Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).

RN [5]

RP CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.



RX MEDLINE=97160848; PubMed=9008168;  
 RA O'Reilly M.S., Behn T., Shing Y., Fukai N., Vassios G., Lane W.S.,  
 RA Flynn E., Birkhead J.R., Olsen B.R., Folkman J.,  
 RT "Endostatin: an endogenous inhibitor of angiogenesis and tumor  
 RT growth.";  
 RL Cell 88:277-285(1997).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF ENDOSTATIN.  
 RA MEDLINE=98169382; PubMed=9501087;  
 RA Honnester E., Sasaki T., Olsen B.R., Timpl R.,  
 RT "Crystal structure of the angiogenesis inhibitor endostatin at 1.5-A  
 RT resolution.";  
 RL EMBO J. 17:1656-1664(1998).  
 CC -|- FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL  
 CC PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY  
 CC BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH  
 CC FACTOR SIGNALLING.  
 CC -|- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM AND A LONG FORM  
 CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -|- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -|- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH  
 CC INTERRUPTED HELICES (FACIT) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, L16898; AAA37434.1; -;  
 DR EMBL, U03714; AAA20657.1; -;  
 DR EMBL, U03715; AAC52901.1; -;  
 DR EMBL, U34606; AAC52901.1; JOINED.  
 DR EMBL, U34608; AAC52901.1; JOINED.  
 DR EMBL, U34609; AAC52901.1; JOINED.  
 DR EMBL, U34610; AAC52901.1; JOINED.  
 DR EMBL, U34611; AAC52901.1; JOINED.  
 DR EMBL, U34612; AAC52901.1; JOINED.  
 DR EMBL, U03716; AAC52901.1; JOINED.  
 DR EMBL, U03718; AAC52901.1; JOINED.  
 DR EMBL, U03715; AAC52902.1; -;  
 DR EMBL, U34607; AAC52902.1; JOINED.  
 DR EMBL, U34608; AAC52902.1; JOINED.  
 DR EMBL, U34609; AAC52902.1; JOINED.  
 DR EMBL, U34610; AAC52902.1; JOINED.  
 DR EMBL, U34611; AAC52902.1; JOINED.  
 DR EMBL, U34612; AAC52902.1; JOINED.  
 DR EMBL, U03716; AAC52902.1; JOINED.  
 DR EMBL, U03718; AAC52902.1; JOINED.  
 DR EMBL, U11636; AAC52178.1; -;  
 DR EMBL, I22545; AAA19787.1; -;  
 DR PDB; 1KOE; 16-FEB-99.  
 DR MGD; MGI:88451; Coll18a1.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR InterPro; IPR003129; TSPN.  
 DR Pfam; PF02391; Collagen; 8.  
 DR Pfam; PF02210; TSPN; 1.  
 DR SMART; SM00282; LamG; 1.  
 DR SMART; SM00210; TSPN; 1.  
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;  
 KW 3D-structure.  
 FT SIGNAL 1 26 POTENTIAL.  
 FT CHAIN 27 1527 COLLAGEN ALPHA 1(XVII) CHAIN.  
 FT CHAIN 1344 1527 ENDOSTATIN.  
 FT DOMAIN 27 538 NONHEMICAL REGION 1 (NC1).  
 FT DOMAIN 539 565 TRIPLE-HELICAL REGION 1 (COL1).

FT	DOMAIN	566	575	NONHEMICAL REGION 2 (NC2).
FT	DOMAIN	576	649	TRIPLE-HELICAL REGION 2 (COL2).
FT	DOMAIN	649	673	NONHEMICAL REGION 3 (NC3).
FT	DOMAIN	673	795	TRIPLE-HELICAL REGION 3 (COL3).
FT	DOMAIN	795	818	NONHEMICAL REGION 4 (NC4).
FT	DOMAIN	818	901	TRIPLE-HELICAL REGION 4 (COL4).
FT	DOMAIN	901	915	NONHEMICAL REGION 5 (NC5).
FT	DOMAIN	915	957	TRIPLE-HELICAL REGION 5 (COL5).
FT	DOMAIN	957	970	NONHEMICAL REGION 6 (NC6).
FT	DOMAIN	970	1043	TRIPLE-HELICAL REGION 6 (COL6).
FT	DOMAIN	1043	1053	NONHEMICAL REGION 7 (NC7).
FT	DOMAIN	1053	1086	TRIPLE-HELICAL REGION 7 (COL7).
FT	DOMAIN	1086	1098	NONHEMICAL REGION 8 (NC8).
FT	DOMAIN	1098	1122	TRIPLE-HELICAL REGION 8 (COL8).
FT	DOMAIN	1122	1123	NONHEMICAL REGION 9 (NC9).
FT	DOMAIN	1123	1181	TRIPLE-HELICAL REGION 9 (COL9).
FT	DOMAIN	1181	1194	NONHEMICAL REGION 10 (NC10).
FT	DOMAIN	1194	1212	TRIPLE-HELICAL REGION 10 (COL10).
FT	DOMAIN	1212	1527	NONHEMICAL REGION 11 (NC11).
FT	CARBOHYD	338	338	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	700	700	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	DISULFID	1376	1536	
FT	DISULFID	1478	1508	
FT	SITE	1104	1106	
FT	VARSPLIC	213	212	CELL ATTACHMENT SITE (POTENTIAL).
FT	VARSPPLIC	213	212	MISSING (IN SHORT ISOFORM).
FT	CONFLICT	900	900	APPTQPPGSGVLRPAPDPF -> MAPRWHLVDL
FT	CONFLICT	947	947	TSVLIVARVSMAL (IN SHORT ISOFORM).
FT	CONFLICT	964	964	P -> F (IN REF. 4).
FT	CONFLICT	1157	1157	A -> R (IN REF. 4).
FT	CONFLICT	1266	1266	R -> P (IN REF. 4).
FT	CONFLICT	1276	1276	P -> L (IN REF. 4).
FT	CONFLICT	1437	1437	L -> F (IN REF. 4).
FT	CONFLICT	1437	1437	L -> V (IN REF. 4).
SO	SEQUENCE	1527 AA; 156008 MM; 9645045AF140B513 CRC64;		

Query Match 7.9%; Score 146.5; DB 1; Length 1527;  
 Best Local Similarity 23.7%; Pred. No. 0.029;  
 Matches 92; Conservative 22; Mismatches 107; Indels 167; Gaps 21;

QY	2	GRPHSGPGVVRVALLLVGLG-LVSGLSLSPVYVNMNSANKRQAGGYLVLPQIGDRDL	60
DB	551	GPP--GPAGPQ-----GPAGVVGSPNSQPV-----	574
QY	61	LCPRAR-PPGHSSPNYEFYKLYLVGAQRCE-----	93
DB	575	--PGAGPFPQPPGPKD-----GTRGRDGEPEGDPEDGRPGDTGQGPFGTGDVG	624
QY	94	-----APPAPNLLTCDRPLDLRFTTKFQSYSPNLMGHERSHHDYIAT	140
DB	625	PKGEKDPGIGRGPFGP-----GPPGSPSRQDKLTFTDME	661
QY	141	SDTRGSLGSLGVCVLTGKMKVLLRVGSPRGGAVPRKP--VSEMPWRDS-GAHSLE	197
DB	662	GSFGSDIESLRG-----PRGFPGPPPPPPVPLGPRGPRFGINGSYA	704
QY	198	PKENLPDPTSNATSRGAEPP--LP-PPSPPAVAGAACGLALLLGVAGAGCANCRRR	254
DB	705	PGPAGP.PGVV-----GREGPPGFPFGPPGPKKEPP-----FGVAGQKGSV-----	747
QY	255	RAPSSRRHGP--GSRGRGSLGGLGGGGMGRBAPGGLGIALGG--GAADPPPCPHY	311
DB	748	----GVGIFPGKSGKDGIDPIGMPEKSLA--GSPGVGPPGPPGPPGPPGPPGAAAF	800
QY	312	EKVSQDYGHPVYIVQ-----DGPQSP	333
DB	801	DDWEGS-GIPLMTTARSDGLGPPGSP	827

RESULT 32  
 CAFF\_RIFPA STANDARD; PRT; 1027 AA.  
 ID CAFF\_RIFPA  
 AC P30754;

DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Fibril-forming collagen alpha chain.  
OS Riffia pachypoda (tube worm).  
OC Eukaryota; Metazoa; Pogonophora; Vestimentifera; Axonobranchia;  
OC Riffidae; Riffidae; Riffia.  
OX NCBI\_TaxId=6426;  
RN 1)  
RP SEQUENCE.  
RX MEDLINE=93130909; PubMed=1483468;  
RA Mann K., Gaill F., Timpl R.;  
RT "Amino-acid sequence and cell-adhesion activity of a fibril-forming  
RT collagen from the tube worm Riffia pachypoda living at deep sea  
RT hydrothermal vents".  
RL Eur. J. Biochem. 210:839-847 (1992).  
RN 12)  
RP SEQUENCE OF 8-45; 525-618 AND 810-882.  
RC TISSUE-Cuticle;  
RX MEDLINE=92015209; PubMed=1920405;  
RA Gaill F., Wiedemann H., Mann K., Kuhn K., Timpl R., Engel U.;  
RT "Molecular characterization of cuticle and interstitial collagens  
RT from worms collected at deep sea hydrothermal vents".  
RL J. Mol. Biol. 221:209-223 (1991).  
CC -1- FUNCTION: FIBRIL-FORMING COLLAGEN.  
CC -1- SUBUNIT: HOMOTETRAMER.  
CC -1- PTM: ALL THE "X" ARE PROBABLY GLYCOSYLATED HYDROXYLISINES.  
DR PIR; S22915; S22915.  
DR InterPro; IPR000087; Collagen.  
DR Pfam; PF01391; Collagen; 16.  
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Glycoprotein; Collagen.  
FT DOMAIN 1 1023 NONHELICAL REGION (N-TERMINAL).  
FT DOMAIN 1024 1027 TRIPLE-HELICAL REGION (C-TERMINAL).  
FT MOD RES 21 21 HYDROXYLATION (PARTIAL).  
FT MOD RES 24 24 HYDROXYLATION (PARTIAL).  
FT MOD RES 27 27 HYDROXYLATION (PARTIAL).  
FT MOD RES 39 39 HYDROXYLATION (PARTIAL).  
FT MOD RES 53 53 HYDROXYLATION (PARTIAL).  
FT MOD RES 54 54 HYDROXYLATION (PARTIAL).  
FT MOD RES 72 72 HYDROXYLATION (PARTIAL).  
FT MOD RES 90 90 HYDROXYLATION (PARTIAL).  
FT MOD RES 93 93 HYDROXYLATION (PARTIAL).  
FT MOD RES 123 123 HYDROXYLATION (PARTIAL).  
FT MOD RES 128 128 HYDROXYLATION (PARTIAL).  
FT MOD RES 150 150 HYDROXYLATION (PARTIAL).  
FT MOD RES 161 161 HYDROXYLATION (PARTIAL).  
FT MOD RES 162 162 HYDROXYLATION (PARTIAL).  
FT MOD RES 164 164 HYDROXYLATION (PARTIAL).  
FT MOD RES 165 165 HYDROXYLATION (PARTIAL).  
FT MOD RES 174 174 HYDROXYLATION (PARTIAL).  
FT MOD RES 177 177 HYDROXYLATION (PARTIAL).  
FT MOD RES 180 180 HYDROXYLATION (PARTIAL).  
FT MOD RES 183 183 HYDROXYLATION (PARTIAL).  
FT MOD RES 207 207 HYDROXYLATION (PARTIAL).  
FT MOD RES 216 216 HYDROXYLATION (PARTIAL).  
FT MOD RES 219 219 HYDROXYLATION (PARTIAL).  
FT MOD RES 228 228 HYDROXYLATION (PARTIAL).  
FT MOD RES 237 237 HYDROXYLATION (PARTIAL).  
FT MOD RES 243 243 HYDROXYLATION (PARTIAL).  
FT MOD RES 249 249 HYDROXYLATION (PARTIAL).  
FT MOD RES 255 255 HYDROXYLATION (PARTIAL).  
FT MOD RES 273 273 HYDROXYLATION (PARTIAL).  
FT MOD RES 276 276 HYDROXYLATION (PARTIAL).  
FT MOD RES 285 285 HYDROXYLATION (PARTIAL).  
FT MOD RES 291 291 HYDROXYLATION (PARTIAL).  
FT MOD RES 303 303 HYDROXYLATION (PARTIAL).  
FT MOD RES 306 306 HYDROXYLATION (PARTIAL).  
FT MOD RES 312 312 HYDROXYLATION (PARTIAL).  
FT MOD RES 321 321 HYDROXYLATION (PARTIAL).  
FT MOD RES 327 327 HYDROXYLATION (PARTIAL).  
FT MOD RES 339 339 HYDROXYLATION (PARTIAL).

FT MOD RES 342 342 HYDROXYLATION (PARTIAL).  
FT MOD RES 348 348 HYDROXYLATION (PARTIAL).  
FT MOD RES 351 351 HYDROXYLATION (PARTIAL).  
FT MOD RES 366 366 HYDROXYLATION (PARTIAL).  
FT MOD RES 372 372 HYDROXYLATION (PARTIAL).  
FT MOD RES 375 375 HYDROXYLATION (PARTIAL).  
FT MOD RES 381 381 HYDROXYLATION (PARTIAL).  
FT MOD RES 387 387 HYDROXYLATION (PARTIAL).  
FT MOD RES 416 416 HYDROXYLATION (PARTIAL).  
FT MOD RES 417 417 HYDROXYLATION (PARTIAL).  
FT MOD RES 423 423 HYDROXYLATION (PARTIAL).  
FT MOD RES 429 429 HYDROXYLATION (PARTIAL).  
FT MOD RES 432 432 HYDROXYLATION (PARTIAL).  
FT MOD RES 433 433 HYDROXYLATION (PARTIAL).  
FT MOD RES 465 465 HYDROXYLATION (PARTIAL).  
FT MOD RES 483 483 HYDROXYLATION (PARTIAL).  
FT MOD RES 500 500 HYDROXYLATION (PARTIAL).  
FT MOD RES 503 503 HYDROXYLATION (PARTIAL).  
FT MOD RES 506 506 HYDROXYLATION (PARTIAL).  
FT MOD RES 513 513 HYDROXYLATION (PARTIAL).  
FT MOD RES 525 525 HYDROXYLATION (PARTIAL).  
FT MOD RES 533 533 HYDROXYLATION (PARTIAL).  
FT MOD RES 536 536 HYDROXYLATION (PARTIAL).  
FT MOD RES 540 540 HYDROXYLATION (PARTIAL).  
FT MOD RES 546 546 HYDROXYLATION (PARTIAL).  
FT MOD RES 551 551 HYDROXYLATION (PARTIAL).  
FT MOD RES 552 552 HYDROXYLATION (PARTIAL).  
FT MOD RES 561 561 HYDROXYLATION (PARTIAL).  
FT MOD RES 603 603 HYDROXYLATION (PARTIAL).  
FT MOD RES 610 610 IMPERFECTION IN THE GAA REPEAT.  
FT MOD RES 621 621 HYDROXYLATION (PARTIAL).  
FT MOD RES 627 627 HYDROXYLATION (PARTIAL).  
FT MOD RES 645 645 HYDROXYLATION (PARTIAL).  
FT MOD RES 647 647 HYDROXYLATION (PARTIAL).  
FT MOD RES 648 648 HYDROXYLATION (PARTIAL).  
FT MOD RES 663 663 HYDROXYLATION (PARTIAL).  
FT MOD RES 708 708 HYDROXYLATION (PARTIAL).  
FT MOD RES 711 711 HYDROXYLATION (PARTIAL).  
FT MOD RES 714 714 HYDROXYLATION (PARTIAL).  
FT MOD RES 717 717 HYDROXYLATION (PARTIAL).  
FT MOD RES 723 723 HYDROXYLATION (PARTIAL).  
FT MOD RES 744 744 HYDROXYLATION (PARTIAL).  
FT MOD RES 759 759 HYDROXYLATION (PARTIAL).  
FT MOD RES 773 773 HYDROXYLATION (PARTIAL).  
FT MOD RES 774 774 HYDROXYLATION (PARTIAL).  
FT MOD RES 783 783 HYDROXYLATION (PARTIAL).  
FT MOD RES 792 792 HYDROXYLATION (PARTIAL).  
FT MOD RES 815 815 HYDROXYLATION (PARTIAL).  
FT MOD RES 816 816 HYDROXYLATION (PARTIAL).  
FT MOD RES 843 843 HYDROXYLATION (PARTIAL).  
FT MOD RES 849 849 HYDROXYLATION (PARTIAL).  
FT MOD RES 855 855 HYDROXYLATION (PARTIAL).  
FT MOD RES 861 861 HYDROXYLATION (PARTIAL).  
FT MOD RES 867 867 HYDROXYLATION (PARTIAL).  
FT MOD RES 888 888 HYDROXYLATION (PARTIAL).  
FT MOD RES 894 894 HYDROXYLATION (PARTIAL).  
FT MOD RES 903 903 HYDROXYLATION (PARTIAL).  
FT MOD RES 915 915 HYDROXYLATION (PARTIAL).  
FT MOD RES 933 933 HYDROXYLATION (PARTIAL).  
FT MOD RES 939 939 HYDROXYLATION (PARTIAL).  
FT MOD RES 945 945 HYDROXYLATION (PARTIAL).  
FT MOD RES 954 954 HYDROXYLATION (PARTIAL).  
FT MOD RES 963 963 HYDROXYLATION (PARTIAL).  
FT MOD RES 966 966 HYDROXYLATION (PARTIAL).  
FT MOD RES 984 984 HYDROXYLATION (PARTIAL).  
FT MOD RES 990 990 HYDROXYLATION (PARTIAL).  
FT MOD RES 1010 1010 HYDROXYLATION (PARTIAL).  
FT MOD RES 1011 1011 HYDROXYLATION (PARTIAL).  
FT MOD RES 1013 1013 HYDROXYLATION (PARTIAL).  
FT MOD RES 1014 1014 HYDROXYLATION (PARTIAL).  
FT MOD RES 1016 1016 HYDROXYLATION (PARTIAL).  
FT MOD RES 1017 1017 HYDROXYLATION (PARTIAL).  
FT MOD RES 1019 1019 HYDROXYLATION (PARTIAL).

MOD RES 1020 1020 HYDROXYLATION.  
 FT VARIANT 903 903 P -> A.  
 SQ SEQUENCE 1027 AA; 94353 MW; 9B043FA7788389C1 CRC64;  
 Query Match 7.8%; Score 146; DB 1; Length 1027;  
 Best Local Similarity 22.2%; Pred. No. 0.021;  
 Matches 82; Conservative 24; Mismatches 113; Indels 150; Gaps 17;

46 GGYLYIPQIDRLICPRAPPPHSPNYEFYKLYVGAQRCE----- 93  
 4 GERYAQAVGP-----IGPRG--PPGPPSPGQO-----GYQGLRKEPPGDSGMPGIGK 50  
 94 -APPAPNLLTCDRPDLRLFTIKQEVSPNLWGHEFRSHHDYIITSPDTRGLESLQ 152  
 51 RGPPEP-----AGIAG-----XSGDDGRGEPGR 75  
 153 GGVCLT-----RGMKYL-----LRVQSFRGQ--AVPRKPYSEM 184  
 76 GGIQMGPRGAGMPGMPGYPXGHRGFRGLSGSGEQKSGNQGPDPGPGAPGPGIPGR 135  
 185 PMERDGAHSLERPEKENTLPDPTSNATSRGABGLPPSPMAVAGAGLALL----- 238  
 136 GGTGSRG-----RDGKSGLP-----LRGVDGLAGPPGPPGIGTSGSPGPPGPGSK 183  
 239 -----LLGVAGAGGAMCMRRRAKPSRHPG--PGSFGSGSLGLGCGGGM----- 283  
 184 GDRGSGSLXGAGGLGPGVGLSGQPGVAGENGHPGMPGMDANGEPGASGESGLGPGSGFP 243  
 284 GPR-----EABPGLGIALRG-----GAADPPCPHYKVSQDYGHVYI 324  
 244 GPRGMPGTAGSPGAGXGDDGPTGEGRPGAPGVXGSSGPP-----GDVAGPCHA 294  
 325 VDDGPPQSP 333  
 295 GEAGRGSP 303

RESULT 33  
 CALH BOVIN STANDARD; PRT; 674 AA.  
 AC P23206;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Collagen alpha 1(X) chain precursor.  
 GN COL10A1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Cartilage;  
 RX MEDLINE=9111311; PubMed=1703407;  
 RA Thomas J.T., Kwan A.P.U., Grant M.E., Boot-Handford R.F.;  
 RT "Isolation of cDNAs encoding the complete sequence of bovine type X  
 collagen. Evidence for the condensed nature of mammalian type X  
 collagen genes";  
 RL Biochem. J. 273:141-148(1991).  
 CC -1- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTROPHIC  
 CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE  
 MINERALIZATION ZONES OF HYALINE CARTILAGE.  
 CC -1- SUBUNIT: HOMOTRIMER.  
 CC -1- PTV: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: STRONG. TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.  
 CC -1- SIMILARITY: CONTAINS 1 C10 DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.ebi-sib.ch/announce/>  
 CC or send an email to [license@ebi-sib.ch](mailto:license@ebi-sib.ch)).  
 CC -----  
 CC EMBL: X53556; CAA37624.1; -  
 CC PIR: S13301; S13301.  
 CC InterPro: IPR001073; C1Q.  
 CC InterPro: IPR000877; Collagen.  
 CC Pfam: PF0386; C1Q; 1.  
 CC PRINTS: PR00007; COMPLEMENTC1Q.  
 CC ProDom: PD00007; Collagen; 1.  
 CC SMART: SMC010; C1Q; 1.  
 CC PROSITE: PS0113; C1Q; 1.  
 CC DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Cartilage; Collagen; Signal; Glycoprotein.  
 FT SIGNAL 1 18  
 FT CHAIN 19 674 COLLAGEN ALPHA 1(X) CHAIN.  
 FT DOMAIN 19 56 NONHELICAL REGION (NC2).  
 FT DOMAIN 57 519 TRIPLE-HELICAL REGION.  
 FT DOMAIN 520 674 NONHELICAL REGION (NC1).  
 FT DOMAIN 539 674 C1Q.  
 FT DISULFID 194 197 BY SIMILARITY.  
 FT MOD RES 460 460 HYDROXYLATION (BY SIMILARITY).  
 FT MOD RES 463 463 HYDROXYLATION (BY SIMILARITY).  
 FT CARBOHYD 611 611 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 674 AA; 65546 MW; CD4CA73A03E004CA CRC64;

Query Match 7.8%; Score 145; DB 1; Length 674;  
 Best Local Similarity 24.2%; Pred. No. 0.016;  
 Matches 100; Conservative 30; Mismatches 125; Indels 158; Gaps 24;

15 ALLILGLVGLVGLSLPEYVNSANKRFQAEQVLYLPQIDRLICPR----- 65  
 6 ALLILMSLVNVLHG-----VFY--TERYQTPRG--IKGPPSNTKQFFTPAIKKGKGLR 56  
 66 -----RPPGP-----HSSPNTEFYKLYLVG-----GAQGR-RCEAPPANLLLT 103  
 57 GEGGIPGPPGAPGPRGHPPS-----GPGKRGTSPPGQPGPLPGPPPS--A 104  
 104 CDRPDLRLFTIKQEVSPNLWGHEFRSHHDYIITADGTRGLESLQ----- 153  
 105 TGRPGI-----PGLPGKQ-----GEB-GLNGKGDIGPAGLGP 137  
 154 -----GVCLTRGMKVLRLVG--QSPRGAVPR-----KPYSEMP--MERDRGAHSL 196  
 138 RGPGRPPGIPGPAISVPEKPPGPPTGEPGRGPGKSGVGLNGQKGMCHCTPC 197  
 197 EPKENTLPDPTSNATSRGABGLPPP-----SMAVAGAGLALLLGVAGAGGA 248  
 198 RPERGLPG-----PQPTGPPGPPGVKKGENGGLPGQPLKGD-----QGVFERGA 245  
 249 MCMRRRAKPSRHP-----GGSFSGGSLGLGCGGGMPEAEPEGLGIA----- 296  
 246 -----APSPGPPGPPGQGBEGIGKPGAPGIPQPGIPGKKGPGAPGTAGLGAPGF 298  
 297 -----LRG-GAADPPCPHYKVSQDYGHVYVDDP-----PQSPENI 336  
 299 GKFGPLGKRGQRPVGLPSPGPAKSGGPPGAGLPGSGNMGPPGPKGI 351

RESULT 34  
 CALH HUMAN STANDARD; PRT; 1516 AA.  
 AC P39060; Q9Y6Q8; Q9Y6Q7; Q9YK38;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 1(XVII) chain precursor [Contains: Endostatin].  
 GN COL18A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.



[illegible]

```

RT      "Clonin and developmental expression of the alpha 3 chain of chicken
RT      type IX collagen."
RL      J. Biol. Chem. 267:10070-10076 (1992).
CC      -1- FUNCTION: COLLAGEN TYPE IX IS A MINOR CARTILAGE NONFIBRILLAR
CC      COLLAGEN. IT IS ASSOCIATED WITH TYPE II COLLAGEN FIBRILS.
CC      -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1 (IX) ,
CC      ALPHA 2 (IX), AND ALPHA 3 (IX).
CC      -1- PM: PROLINS ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC      UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M83179; AAB59960.1; -.
DR      PIR; S22918; S22918.
DR      InterPro; IPR000087; Collagen.
DR      Pfam; PF01391; Collagen; 11.
DR      ProDom; PD000007; Collagen; 1.
DR      Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
DR      Glycoprotein; Cartilage; Collagen; Signal.
DR      SIGNAL
DR      CHAIN
FT      CHAIN
FT      DOMAIN
FT      DOMAIN
FT      DOMAIN
FT      DOMAIN
FT      DOMAIN
FT      SITE
FT      SITE
FT      CARBOHYD
SQ      SEQUENCE
Query Match
Best Local Similarity
Matches
64; Conservative
7.8%; Score 144; DB 1; Length 675;
28.8%; Pred. No. 0.018;
9; Mismatches 71; Indels 78; Gaps 12;
166 RVGGSPRGGAIVPRKVSMPMERDRGAHSLPEKGEN;PGDPTSVATSGAAGPLPPSM 225
23 RVG-PQGPDPGRPGSGKDIDG---EPGSGSLGPP------GPKGATGX 65
226 PAVAGAG-GLATLLLVAGAGAMCWRRRANPSSRHQ-----PQSFGRG- 272
66 PQAAGEAGLPGHP-----GVDDGLTG------DPPGPGMPGDRGALGPAGPPGAGKGL 115
273 -----GSLGGGGGGM-----GPR-----EALPGELGIALNG- 300
116 PGPPEPPPGSGLPGSNGRGRPPPGSLGFGPPPPPPPPPLGALILPBGAGLQCPALCP 175
301 -GAADPPFCPHYEKVSGDYGHFVYIVQDP-----PQSPENI 336
176 PGPPEPPGMPGPKGHTGKHGGGBELGKGGKSGPSPPPPGI 217
CA19_HUMAN
CA19_HUMAN STANDARD; PRT; 921 AA.
P20849; Q9Y6P2; Q9Y6P3; Q9H151; Q9H152; Q992255; Q13699; Q13700;
01-EB-1991 (Rel. 17, Created)
15-JUN-2002 (Rel. 41, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
Collagen alpha 1(IX) chain precursor.
COL9A1.
Homo sapiens (Human) .
Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]

```



Query Match 7.8%; Score 144; DB 1; Length 921;  
 Best Local Similarity 22.3%; Pred. No. 0.025;  
 Matches 84; Conservative 21; Mismatches 120; Indels 152; Gaps 13;

CC 2 GPHSGGPGVRRGALLLGLVGLVGLSLPEPYMSANKRQAEQGYVLPQIGRDL 61  
 DB 391 GPP--GPPPPR-----GTIGFHDG-----DPL 410

CC 62 CRRAPPGRPHSSPNYEFKLYIVG-----GAQRRRCAPRPNLLITCDRDLRFTIK 116  
 DB 411 CRRAPPGRPHSSPNYEFKLYIVG-----GAQRRRCAPRPNLLITCDRDLRFTIK 116

CC 117 PCEYBNLMGHEFRSHHDYIANTSDGTREGESLQGVCLTRGKMLRYVQSPRGSAV 176  
 DB 441 -----PGRGCH-----KGEEDQG-----ELGEGAGCPGPAQ 468

CC 177 PRKPYSEMPMERDRGAHSL--EPGKENLPDPTSNAT-----SRGAGPLPFP 223  
 DB 469 GURGITGVGDKGKAGLDEPQPGI.PGAPDQGRGPPGKAGPDRGAGARGIF 528

CC 224 SMPAVAGAGGALLLLGVAGAGAMCRRRAKSESRRHG-----PSFRGSGSLG 276  
 DB 529 GLPGRGDTG-----LPVDGRDGI.PGMPGTGEPKPGPPEDAGLQGLPEVPGIPGAKG 583

CC 277 LGGGGGMPREAPGELGIALRG-----GAADPPFCPHYEKVSG 316  
 DB 584 VAGEKGSITAPKPGQMGNSKRGQGGPPGVEGPRPGQLPGRSRELTPVSGPLPKLG 643

CC 317 DYGHPIYIVQDGPPOSP 333  
 DB 644 SLGSPGLPGLPGPGLP 660

RESULT 37  
 CHICK CHICK STANDARD; PRT; 674 AA.

AC P08125;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Collagen.alpha.1(X) chain precursor.  
 GN COL10A1.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archaeopteryx; Aves; Neognathae; Galliformes; Phasianidae;  
 OC Gallus.  
 OC NCBI\_Taxid=9031;  
 RN 111  
 RN SEQUENCE OF 46-674 FROM N.A., AND SEQUENCE OF 103-117 AND 453-466.  
 RP MEDLINE=86168227; PubMed=3082876;  
 RA Nimmiya Y., Gordon M., van der Rest M., Schmid T., Linsemayer T.,  
 RA Olsen B.R.;  
 RT "The developmentally regulated type X collagen gene contains a long  
 RT open reading frame without introns."  
 RL J. Biol. Chem. 261:5041-5050(1986).  
 RN 121  
 RN SEQUENCE OF 1-75 FROM N.A.  
 RP MEDLINE=89054019; PubMed=2461368;  
 RA Luvall P., Nimmiya Y., Rosenblum N.D., Olsen B.R.;  
 RT "The type X collagen gene. Intron sequences split the 5'-untranslated  
 RT region and separate the coding regions for the non-collagenous amino-  
 RT terminal and triple-helical domains."  
 RL J. Biol. Chem. 263:18378-18385(1988).  
 RN 131  
 RN REVISIONS TO C-TERMINUS.  
 RP MEDLINE=89380199; PubMed=2476437;  
 RA Yamasuchi N., Benya P.D., van der Rest M., Nimmiya Y.;  
 RT "The cloning and sequencing of alpha 1(VIII) collagen cDNAs  
 RT demonstrate that type VIII collagen is a short chain collagen and  
 RT contains triple-helical and carboxyl-terminal non-triple-helical  
 RT domains similar to those of type X collagen."  
 RL J. Biol. Chem. 264:16022-16029(1989).  
 CC -!- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTHROPHIC

CC CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE  
 CC MINERALIZATION ZONES OF HYALINE CARTILAGE.  
 CC -!- SUBUNIT: HOMOTRIMER.  
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -!- SIMILARITY: STRONG, TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.  
 CC -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M13496; AAA48736.1; ALT\_SEQ.  
 CC EMBL; J04194; AAA48634.1; -.  
 CC PIR; A31896; A31896.  
 CC InterPro; IPR001073; C1q.  
 CC InterPro; IPR000087; Collagen.  
 CC Pfam; PF00386; C1q; 1.  
 CC Pfam; PF01391; Collagen; 8.  
 CC PRINTS; PR00007; COMPLEMENTC1Q.  
 CC SMART; SM00110; C1Q; 1.  
 CC PROSITE; PS01113; C1Q; 1.  
 CC DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 CC KW Cartilage; Collagen; Signal.  
 CC  
 CC SIGNAL 1 18  
 CC CHAIN 19 674 COLLAGEN ALPHA 1(X) CHAIN.  
 CC FT 19 52 NONHELICAL REGION (NC2).  
 CC FT DOMAIN 53 512 TRIPLE-HELICAL REGION.  
 CC FT DOMAIN 513 674 NONHELICAL REGION (NC1).  
 CC FT DOMAIN 539 674 C1Q.  
 CC FT MOD\_RES 453 453 HYDROXYLATION.  
 CC FT MOD\_RES 456 456 HYDROXYLATION.  
 CC FT MOD\_RES 456 456 HYDROXYLATION.  
 CC SQ SEQUENCE 674 AA; 66434 MW; EAB48B1EF74B145 CRC64;

Query Match 7.7%; Score 143; DB 1; Length 674;  
 Best Local Similarity 25.0%; Pred. No. 0.021;  
 Matches 77; Conservative 18; Mismatches 91; Indels 122; Gaps 16;

CC 67 PRGPHSSPNYEFY-KLYIVG-----GAQRRRCAPRPNLLITCDRDLRFTIK 117  
 DB 86 PRGPHSSPNYEFY-KLYIVG-----GAQRRRCAPRPNLLITCDRDLRFTIK 117

CC 118 QEYSPNLMGHEFRSHHDYIANTSDGTREGESLQGVCLTRGKMLRYVQSPRGSAV 171  
 DB 126 -----PGRGCH-----KGEEDQG-----ELGEGAGCPGPAQ 468

CC 172 RGAVPRKPYSEMPMERDRGAHSLPEPKENLPDPTSNAT-----SRGAG 218  
 DB 157 RGAVPRKPYSEMPMERDRGAHSLPEPKENLPDPTSNAT-----SRGAG 218

CC 219 PLPPSPAVAGAGGALLLLGVAGAGAMCRRRAKSESRRHG-----PSFRGSGSLG 276  
 DB 200 PLPPSPAVAGAGGALLLLGVAGAGAMCRRRAKSESRRHG-----PSFRGSGSLG 276

CC 275 LGLGGGMPREAPGELGIALRG-----GAADPPFCPHYEKVSG 316  
 DB 254 LGLGGGMPREAPGELGIALRG-----GAADPPFCPHYEKVSG 316

CC 326 QDGPPOSP 333  
 DB 306 QDGPPOSP 333

RESULT 38  
 CA25 HUMAN STANDARD; PRT; 1496 AA.  
 AC P05957;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 15-JUN-2002 (Rel. 41, last annotation update)  
DE Collagen alpha 2 (V) chain precursor.  
CN COL5A2.  
CS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE OF 1-463 FROM N.A.  
RX MEDLINE=69123368; PubMed=2914297;  
RA Woodbury D., Benson-Chanda V., Ramirez F.,  
RT "Amino-terminal propeptide of human pro-alpha 2 (V) collagen conforms  
RL to the structural criteria of a fibrillar procollagen molecule.";  
RL J. Biol. Chem. 264:2735-2738 (1989).  
RN [2]  
RP SEQUENCE OF 398-1496 FROM N.A.  
RX MEDLINE=8714631; PubMed=3029669;  
RA Weil D., Bernard M.P., Gargano S., Ramirez F.,  
RT "The pro alpha 2 (V) collagen gene is evolutionarily related to the  
RL major fibrillar-forming collagens.";  
RN Nucleic Acids Res. 15:181-198 (1987).  
RN [3]  
RP SEQUENCE OF 1227-1496 FROM N.A.  
RX MEDLINE=85289337; PubMed=2411731;  
RA Myers J.C., Ioldi H.R., Seyer J.M., Dion A.S.,  
RT "Complete primary structure of the human alpha 2 type V procollagen  
RL COOH-terminal propeptide.";  
RL J. Biol. Chem. 260:11216-11222 (1985).  
RN [4]  
RP SEQUENCE OF 1449-1496 FROM N.A.  
RX MEDLINE=89138450; PubMed=3224983;  
RA Tsipouras P., Schwartz R.C., Hiddell A.C., Salkeid C.S., Weil D.,  
RA Ramirez F.,  
RT "Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2,  
RL located on the long arm of human chromosome 2.";  
RL Genomics 3:275-277 (1988).  
RN [5]  
RP SEQUENCE OF 208-227.  
RX TISSUE=Placenta;  
RC MEDLINE=92239022; PubMed=1571108;  
RA Mann K.,  
RT "Isolation of the alpha 3-chain of human type V collagen and  
RL characterization by partial sequencing.";  
RL Biol. Chem. Hoppe-Seyler 373:65-75 (1992).  
RN [6]  
RP SEQUENCE OF 288-297 AND 606-617.  
RX TISSUE=Bone;  
RC MEDLINE=94237164; PubMed=8181482;  
RA Moradi-Arneil M., Rousseau J.C., Kleman J.P., Champliand M.F.,  
RA Boutillon M.M., Bernillon J., Wallach J.M., van der Reef M.,  
RT "Diversity in the processing events at the N-terminus of type-V  
RL collagen.";  
RL Eur. J. Biochem. 221:987-995 (1994).  
RN [7]  
RP DISEASE.  
RX PubMed=9425231;  
RA Michalickova K., Susic M., Willing M.C., Wensstrup R.J., Cole W.G.,  
RT "Mutations of the alpha2 (V) chain of type V collagen impair matrix  
RL assembly and produce Ehlers-Danlos syndrome type I.";  
RL Hum. Mol. Genet. 7:249-255 (1998).  
RN [8]  
RP VARIANT EDS-II ARG-960.  
RX PubMed=9783710;  
RA Burrows N.P., Martin S., Nicholls A.C., Harrison J.B., Pope F.M.,  
RA Burrows N.P.,  
RT "A single base mutation in COL5A2 causes Ehlers-Danlos syndrome type  
RL II.";  
RL J. Med. Genet. 35:846-848 (1998).  
CC -1- FUNCTION: TYPE V COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
CC (FIBRILLAR FORMING COLLAGEN). IT IS A MINOR CONNECTIVE TISSUE  
CC COMPONENT OF NEARLY UBQUITOUS DISTRIBUTION. TYPE V COLLAGEN BINDS  
CC TO DNA, HEPARAN SULFATE, THROMBOSPONDIN, HEPARIN, AND INSULIN.  
CC -1- SUBUNIT: TRIMERS OF TWO ALPHA 1 (V) AND ONE ALPHA 2 (V) CHAINS IN

CC MOST TISSUES AND TRIMERS OF ONE ALPHA 1(V), ONE ALPHA 2(V), AND  
CC ONE ALPHA 3(V) CHAINS IN PLACENTA.  
CC -I- PPM: PROLINS AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
CC -I- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos syndrome  
CC graveis or type I (EDS-I), and Ehlers-Danlos syndrome mlts or type  
CC II (EDS-II). Ehlers-Danlos syndrome is a genetically and  
CC phenotypically heterogeneous connective-tissue disorder  
CC characterized by loose-jointedness and fragile, velvety,  
CC stretchable, bruisable skin that heals with peculiar 'cigarette-  
CC paper' scars. EDS-I and EDS-II are autosomal dominant traits.  
CC -I- SIMILARITY: CONTAINS 1 VWFC DOMAIN.

---

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.ebi.ac.uk/termsandconditions/>  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

---

DR EMBL; J044758; AAA51859.1; -;  
DR EMBL; X04758; CAA28454.1; -;  
DR EMBL; M1718; AA52058.1; -;  
DR PIR; A25374; A25374.  
DR PIR; A25874; A25874.  
DR PIR; A30017; A30017.  
DR PIR; A31427; A31427.  
DR GeneW; HGNC:2210; COL5A2.  
DR MIM; 120190; -;  
DR MIM; 130000; -;  
DR DR InterPro; IPRO00087; Collagen.  
DR DR InterPro; IPRO00885; Fib collagen\_C.  
DR DR InterPro; IPRO01007; VWF\_C.  
DR Pfam; PF00391; WVC; 1.  
DR Pfam; PF01391; Collagen; 18.  
DR Pfam; PF01410; COLFR; 1.  
DR ProdDom; PD000007; Collagen; 5.  
DR ProdDom; PD002078; Fib collagen\_C; 1.  
DR SMART; SMO0038; COLFI; 1.  
DR SMART; SMO0214; WVC; 1.  
DR PROSITE; PS01208; VWFC; 1.  
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Glycoprotein; Collagen; Signal; Ehlers-Danlos syndrome;  
KW Disease mutation.

KW FT SIGNAL	1	26
KW FT CHAIN	27	1226
KW FT PROPEP	1227	1496
FT FT DOMAIN	39	97
FT FT MOD_RES	290	290
FT FT MOD_RES	293	293
FT FT MOD_RES	296	296
FT FT MOD_RES	608	608
FT FT MOD_RES	614	614
FT FT VARIANT	960	960

FT FT CONFLICT 292 292 A -> P (IN REF. 6).  
FT FT CONFLICT 1418 1418 K -> T (IN REF. 3).  
FT FT CONFLICT 1438 1438 F -> S (IN REF. 3).  
FT FT CONFLICT 1460 1460 E -> Q (IN REF. 4).  
FT FT CONFLICT 1496 1496 V -> A (IN REF. 4).  
SQ SEQUENCE 1496 AA; 144720 MW; 82827C17A8644F5A CRC64;

Query Match 7.7%; Score 142.5; DB 1; Length 1496;  
Best Local Similarity 29.18; Pred. No. 0.051; Idels 43; Gaps 9;  
Matches 55; Conservative 15; Mismatches 76;

```

Oy      169 QSPRG--GAV--PRPVSEMPWERDRGAHSLEPKENLPG-----DPTSNTSRGAE 217
           :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      500 RGRPDDEPTLLRPGGVYVGRGAPGNRF-----PSDDGLPGKKAQGGERGYPVGSSGPKKSQ 554
           ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Oy      218 GPUPPPNPANVAAGAGLALLLVNVAAGAMCMRRRAKSESRRHPD---GSFGNG 273
           ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```





RX MEDLINE=91226374; PubMed=1709252;  
 RA Constantino C.D., Jimenez S.A.;  
 RT "Structure of CDNA encoding the triple-helical domain of murine  
 RT alpha 2 (VI) collagen chain and comparison to human and chick  
 RT homologues. Use of polymerase chain reaction and partially degenerate  
 RT oligonucleotide for generation of novel cDNA clones.";  
 RL Matrix 11:1-9(1991).  
 RN [4]  
 RP SEQUENCE OF 659-1029 FROM N.A.  
 RX MEDLINE=9326888; PubMed=849506;  
 RA Zhang R.Z., Fan T.C., Timpl R., Chu M.-L.;  
 RT "Cloning and sequence analysis of cDNAs encoding the alpha 1, alpha 2  
 RT and alpha 3 chains of mouse collagen VI.";  
 RL Biochem. J. 291:787-792(1993).  
 CC -1- FUNCTION: COLLAGEN VI ACTS AS A CELL-BINDING PROTEIN.  
 CC -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1 (VI),  
 CC ALPHA 2 (VI), AND ALPHA 3 (VI).  
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADIPOSE TISSUE, LUNG,  
 CC ADRENAL GLANDS AND OVARY. LOWER LEVELS IN TESTIS, TONGUE, SKIN,  
 CC KIDNEY, HEART, INTESTINE AND SPLEEN. NO EXPRESSION IN SKELETAL  
 CC MUSCLE OR LIVER.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC -1- UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: CONTAINS 3 WFPA DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X65582; CAA46541.1; -  
 DR EMBL; X62332; CAA44206.1; -  
 DR EMBL; Z18272; CAA79153.1; -  
 DR EMBL; L06343; AAA37441.1; -  
 DR PIR; S13745; S13745.  
 DR MGD; MG1:88460; Col6a2.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF01381; Collagen; 4.  
 DR PRINTS; PR00453; VWFADOMAIN.  
 DR SMART; SM00327; VWA; 3.  
 DR PROSITE; PS50234; VWFA; 3.  
 DR KMW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KM Glycoprotein; Cell adhesion; Collagen; Signal.  
 FT SIGNAL 1 20  
 FT CHAIN 21 1029  
 FT DOMAIN 21 265  
 FT DOMAIN 266 600  
 FT DOMAIN 601 1029  
 FT DOMAIN 54 178  
 FT DOMAIN 623 748  
 FT DOMAIN 843 965  
 FT SITE 376 378  
 FT SITE 436 438  
 FT SITE 499 501  
 FT SITE 508 510  
 FT SITE 549 551  
 FT CARBOHYD 150 150  
 FT CARBOHYD 337 337  
 FT CARBOHYD 640 640  
 FT CARBOHYD 795 795  
 FT CARBOHYD 907 907  
 FT CARBOHYD 963 963  
 FT CONFLICT 268 268  
 FT CONFLICT 804 804  
 FT CONFLICT 848 848  
 FT CONFLICT 962 966  
 FT CONFLICT 976 977  
 SQ SEQUENCE 1029 AA; 109811 MW; EDALBA0AF266FSD CRC64;

Query Match 7.78; Score 142; DB 1; Length 1029;  
 Best Local Similarity 29.78; Pred. No. 0.038;  
 Matches 90; Conservative 16; Mismatches 125; Indels 72; Gaps 20;  
 QY 68 PGPSPVYERFYKLYLVGAGRCCEAPAPNLLTCDPRLDRTFTKFOEYSPNLMGH 127  
 DB 265 PGPSPVYERFYKLYLVGAGRCCEAPAPNLLTCDPRLDRTFTKFOEYSPNLMGH 318  
 QY 128 --EPSSHVDYITINSDGTREGLSLOG--GVCITRKMKVLL-RVG-QSPRGCAVPRKE- 180  
 DB 319 KGEF-----GSDG-RKAGAGLAKNG--TDCQKGLGIGPPGCKGDDGSRGPD 364  
 QY 181 ----VSEMPMER-DRGA-ASHLEPGKENVLPDPTSNATSGAEPFLPPSPAVAGAAG 234  
 DB 365 GYPEAGPGSRGQAGKDBSGRGRGPPDP--GDGSKGYQANNAPGSPYKGGKGG 423  
 QY 235 LALLLVGAGAGAMCWRRRRAKP-----SESHPGP---GSFRCGSTGLG 278  
 DB 424 -----PGRGPKEPGRGDPGTGKGPQSDGPKXKEDPGEGPGLAGEVSGKAK 475  
 QY 279 GGGGW-GPRELPEGLIALRGGAADPPCPHYEKVSGYGHVYIVODGPPSP--PN 335  
 DB 476 GDRGPGRGPQ-GALBPGKQSGSRGP-----GAGRGDSGQPPGKDPGRPG 524  
 QY 336 IYV 338  
 DB 525 FSX 527

## RESULT 41

CA24\_ASCSU STANDARD; PRT; 1763 AA.

AC P27393;  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DB Collagen alpha 2 (IV) chain precursor.  
 OS Ascaris suum (pig roundworm) (Ascaris lumbricoides).  
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascarididae;  
 CC Ascarididae; Ascaris.  
 CC NCBI\_TaxId=6253;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91340768; PubMed=1714907;  
 RA Pettit J., Kingston I.B.;  
 RT "The complete primary structure of a nematode alpha 2(IV) collagen  
 RT and the partial structural organization of its gene.";  
 RL J. Biol. Chem. 266:16149-16156(1991).  
 CC -1- FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES.  
 CC -1- SUBUNIT: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.  
 CC TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH  
 CC INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NC1  
 CC DOMAINS.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND II; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS  
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERUPTIONS OF THE  
 CC G-X-Y REPEATS IN THE LONG CENTRAL, TRIPLE-HELICAL DOMAIN (WHICH MAY  
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL  
 CC TRIPLE-HELICAL 7S DOMAIN.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF  
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE  
 CC IV COLLAGENS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M67207; AAA18014.1; -

DR PIR; S16366; S16366.

DR InterPro; IPR000087; Collagen.

DR InterPro; IPR001442; ProcollagenC4.

DR Pfam; PF01391; Collagen; 25.

DR Pfam; PF01413; C4; 2.

DR ProDom; PD000007; Collagen; 3.

DR ProDom; PD003923; ProcollagenC4; 2.

DR SMART; SW00111; C4; 2.

KW Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;

KW Alternative splicing; Glycoprotein; Signal.

FT SIGNAL 1 26

FT CHAIN 1 26

FT DOMAIN 27 1763

FT DOMAIN 27 42

FT DOMAIN 43 1529

FT DOMAIN 1530 1763

FT DISULFID 1548 1637

FT DISULFID 1561 1634

FT DISULFID 1593 1599

FT DISULFID 1566 1752

FT DISULFID 1680 1749

FT DISULFID 1702 1709

FT CARBOHYD 126 126

FT CARBOHYD 249 249

FT VARSPLIC 230 266

SQ SEQUENCE 1763 AA; 168526 MW; 3044528BC06AAE0D CRC64;

Query Match 7.7%; Score 142; DB 1; Length 1763;

Best Local Similarity 25.5%; Pred. No. 0.064;

Matches 97; Conservative 24; Mismatches 127; Indels 132; Gaps 23;

QY 2 GPHSGPGGVAVGALLLGLVGLV-----VSGLSLEP-----VYNSANKRFOAEQGYLY 51

DB 82 GPP--GPGGK--GDRGIIIGVPGFPNDANGPRGPBPAGAPMGCGCTGADG---V 134

QY 52 PIGGRDLICGRARP--RGRSSPNVEFYKLYLVGAGGRCEA--PPANLLLTCDRP 107

DB 135 P-----GLPGRPMGPRGPVCGMKGBPAIGYAGAPGKGDGMFMGL----- 181

QY 108 DLDLRFITKFOEYSPNMGHEFRSHDYIATSDTREGLESLOGGVCLTRGKVLRY 167

DB 182 -----PGRPG-----RDGPRGKGD---DRG-----DV 200

QY 166 GGS-PRG-----GAV-----PRKPVSEMPMERDRGAASLE--P 198

DB 201 GAGRGPRGPGGAPGNPGIGSIGKPGDGEQPRGPGPAPVASTAKGTITIGPEGAP 260

QY 199 GENIIPGPTSNATRGAGELPPSPMAVAGAGAGLLLLLVGAGAGACMRRRAKP 258

DB 261 GAKGKGRDP-----GAGRGPRGPTGVAQGP-----LPKMGKGLSGPRAPRKE 308

QY 259 SESRHPGPGSGFGRGSLGLGGGGMGPRAEPABEGELIALRG--GAADPPCPHYEKVSG 316

DB 309 GRPGPGPGPGF--KGRDGLDGLPVPGLPGQKGEAGFGDRGAKARPGPPG---GG 362

QY 317 DYGHFVYIVQDGPSPSPNI 336

DB 363 EF-----SDGFP--GPFEL 374

RESULT 42

CA13 HUMAN

AC P02461; O15112; STANDARD; PRT; 1466 AA.

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Collagen alpha 1(III) chain precursor.

OS COL3A1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo.

OX NCBI\_Taxid:9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-Skin fibroblast;

RX MEDLINE=89350838; PubMed=2764886;

RA Ala-Kokko U., Kontusaari S., Baldwin C.T., Kuivaniemi H.,

RA Prockop D.J.;

RT "Structure of cDNA clones coding for the entire propeptide alpha 1 (III)

RT chain of human type III procollagen. Differences in protein structure

RT from type I procollagen and conservation of codon preferences.";

RL Biochem. J. 260:509-516(1989).

RN [2]

RP SEQUENCE OF 149-1225 FROM N.A.

RX MEDLINE=89386015; PubMed=2780304;

RA Janeczko R.A., Ramirez F.;

RT "Nucleotide and amino acid sequences of the entire human alpha 1

RT (III) collagen.";

RL Nucleic Acids Res. 17:6742-6742(1989).

RN [3]

RP SEQUENCE OF 168-398.

RX MEDLINE=77134724; PubMed=557335;

RA Seyer J.M., Kang A.H.;

RT "Covalent structure of collagen: amino acid sequence of cyanogen

RT bromide peptides from the amino-terminal segment of type III collagen

RT of human liver.";

RL Biochemistry 16:1158-1164(1977).

RN [4]

RP REVISIONS.

RA Seyer J.M.;

RL Submitted (DEC-1977) to the PIR data bank.

RN [5]

RP SEQUENCE OF 399-727.

RX MEDLINE=79000343; PubMed=687591;

RA Seyer J.M., Kang A.H.;

RT "Covalent structure of collagen: amino acid sequence of five

RT consecutive CNBR peptides from type III collagen of human liver.";

RL Biochemistry 17:3404-3411(1978).

RN [6]

RP SEQUENCE OF 728-964.

RX MEDLINE=80198282; PubMed=6246925;

RA Seyer J.M., Mainardi C., Kang A.H.;

RT "Covalent structure of collagen: amino acid sequence of alpha 1

RT (III)-CB5 from type III collagen of human liver.";

RL Biochemistry 19:1583-1589(1980).

RN [7]

RP SEQUENCE OF 950-1466 FROM N.A.

RX MEDLINE=8189827; PubMed=3357782;

RA Mankoo B.S., Dalglish R.;

RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";

RL Nucleic Acids Res. 16:2337-2337(1988).

RN [8]

RP REVISION TO 1184.

RX MEDLINE=8908346; PubMed=3211760;

RA Molyneux K., Dalglish R.;

RT "Human type III collagen 'variant' is a cDNA cloning artefact.";

RL Nucleic Acids Res. 16:11833-11833(1988).

RN [9]

RP SEQUENCE OF 1065-1466 FROM N.A.

RX MEDLINE=85087944; PubMed=6096827;

RA Loidl H.R., Brinker U.M., May M., Pihlajaniemi T., Morrow S.,

RA Rosenbloom J., Myers U.C.;

RT "Molecular cloning and carboxyl-propeptide analysis of human type III

RT procollagen.";

RL Nucleic Acids Res. 12:9383-9394(1984).

RN [10]

RP SEQUENCE OF 965-1200.

RX MEDLINE=81208139; PubMed=7016180;

RA Seyer J.M., Kang A.H.;

RT "Covalent structure of collagen: amino acid sequence of alpha  
RT 1(III)-OB9 from type III collagen of human liver.";  
RL Biochemistry 20:2621-2627(1981).  
RN [11]  
RP SEQUENCE OF 1176-1466 FROM N.A.  
RX MEDLINE=85157600; PubMed=2579949;  
RA Chu M.-L., Weil D., de Wet M.J., Bernard M.P., Sippola M., Ramirez F.;  
RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1  
RT (III) collagen. Partial characterization of the 3' end region of the  
RT gene.";  
RL J. Biol. Chem. 260:4357-4363(1985).  
RN [12]  
RP SEQUENCE OF 1161-1200 FROM N.A.  
RX MEDLINE=86187804; PubMed=3754452;  
RA Miskulin P., Dalgleish R., Kluge-Beckerman B., Renard S.I.,  
RA Tolosiev P., Brantly M., Crystal R.G.;  
RT "Human type III collagen gene expression is coordinately modulated  
RT with the type I collagen genes during fibroblast growth.";  
RL Biochemistry 25:1408-1413(1986).  
RN [13]  
RP SEQUENCE OF 1-170 FROM N.A.  
RC TISUE-Placenta;  
RX MEDLINE=88303360; PubMed=3405773;  
RA Toman D., Ricca G., de Crombrughe B.;  
RT "Nucleotide sequence of a cDNA coding for the amino-terminal region  
RT of human prepro alpha 1(III) collagen.";  
RL Nucleic Acids Res. 16:7201-7201(1988).  
RN [14]  
RP SEQUENCE OF 1-176 FROM N.A.  
RX MEDLINE=89378752; PubMed=2777083;  
RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;  
RT "Cloning and analysis of the 5' portion of the human type-III  
RT procollagen gene (COL3A1).";  
RL Gene 78:255-265(1989).  
RN [15]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=97255959; PubMed=9101290;  
RA Kuivaniemi H., Tromp G., Prockop D.J.;  
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
RT associated collagen (type IX), and network-forming collagen (type X)  
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
RL Hum. Mutat. 9:300-315(1997).  
RN [16]  
RP VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.  
RX MEDLINE=93293989; PubMed=8514866;  
RA Tromp G., Wu Y., Prockop D.J., Madhathari S.L., Kleiner C.,  
RA Barley J.J., Zhuang J., Noertgaard O., Darling R.C., Abbott W.M.,  
RA Cole C.W., Taakkola P., Rytanen M., Pearce W.H., Yao J.S.T.,  
RA Jackson C.E., Michels S.N., Gatalica Z., Ferrell R.B., Umenez S.A.,  
RA Macken C.E., Michaels V.V., Kaya M., Kuivaniemi H.;  
RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations  
RT in the triple-helical domain of type III procollagen are an  
RT infrequent cause of aortic aneurysms.";  
RL J. Clin. Invest. 91:2539-2545(1993).  
RN [17]  
RP VARIANT THR-698.  
RX MEDLINE=91045136; PubMed=2235526;  
RA Zafarullah K., Kleiner C., Tromp G., Kuivaniemi H., Kontusaari S.,  
RA Wu Y., Ganguly A., Prockop D.J.;  
RT "A polymorphism in exon 31 of the COL3A1 gene.";  
RL Nucleic Acids Res. 18:6180-6180(1990).  
RN [18]  
RP VARIANT AORTIC ANEURYSM ARG-786.  
RX MEDLINE=91056145; PubMed=2243125;  
RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;  
RT "A mutation in the gene for type III procollagen (COL3A1) in a family  
RT with aortic aneurysms.";  
RL J. Clin. Invest. 86:1465-1473(1990).  
RN [19]  
RP VARIANT EDS-IV ARG-828.  
RX MEDLINE=94016385; PubMed=8411057;  
RA Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;  
RT "The substitution of glycine 661 by arginine in type III collagen

RT produces mutant molecules with different thermal stabilities and  
RT causes Ehlers-Danlos syndrome type IV.";  
RL J. Med. Genet. 30:690-693(1993).  
RN [20]  
RP VARIANT EDS-IV SER-957.  
RX MEDLINE=89109135; PubMed=2492273;  
RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;  
RT "A single base mutation that substitutes serine for glycine 790 of  
RT the alpha 1 (III) chain of type III procollagen exposes an arginine  
RT and causes Ehlers-Danlos syndrome IV.";  
RL J. Biol. Chem. 264:1349-1352(1989).  
RN [21]  
RP VARIANT EDS-IV VAL-960.  
RX MEDLINE=95266429; PubMed=7749417;  
RA Tromp G., de Paeppe A., Nuytink L., Madhathari S.L., Kuivaniemi H.;  
RT "Substitution of valine for glycine 793 in type III procollagen in  
RT Ehlers-Danlos syndrome type IV.";  
RL Hum. Mutat. 5:179-181(1995).  
RN [22]  
RP VARIANT EDS-IV GLU-1014.  
RX MEDLINE=92316511; PubMed=1352273;  
RA Richards A.V., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,  
RA Pope F.M.;  
RT "A single base mutation in the gene for type III collagen (COL3A1)  
RT converts glycine 847 to glutamic acid in a family with Ehlers-Danlos  
RT syndrome type IV. An unaffected family member is mosaic for the  
RT mutation.";  
RL Hum. Genet. 89:414-418(1992).  
RN [23]  
RP VARIANT EDS-IV ASP-1050.  
RX MEDLINE=90037070; PubMed=2808425;  
RA Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;  
RT "Single base mutation in the type III procollagen gene that converts  
RT the codon for glycine 883 to aspartate in a mild variant of  
RT Ehlers-Danlos syndrome IV.";  
RL J. Biol. Chem. 264:19313-19317(1989).  
RN [24]  
RP VARIANT EDS-IV VAL-1077.  
RX MEDLINE=91374480; PubMed=1895316;  
RA Richards A.J., Lloyd J.C., Ward P.N., de Paeppe A., Narcisi P.,  
RA Pope F.M.;  
RT "Characterization of a glycine to valine substitution at amino acid  
RT position 910 of the triple helical region of type III collagen in a  
RT patient with Ehlers-Danlos syndrome type IV.";  
RL J. Med. Genet. 28:458-463(1991).  
RN [25]  
RP VARIANT EDS-IV GLU-1173.  
RX MEDLINE=93022543; PubMed=1357232;  
RA Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;  
RT Query Match 7.6%; Score 141.5; DB 1; Length 1466;  
RT Best Local Similarity 27.6%; Pred. No. 0.058;  
RT Matches 61; Conservative 13; Mismatches 74; Indels 73; Gaps 10;  
OY 171 PRGAVPRKPVSEMPNERDCA-----AHSLEPKENLPDPTSNAT-----SRAGG 218  
DB 844 PRGSGPAGPFGQVGVGERSGPGAGPFGARGLPFGSGGNDPPGSGSPKDG 903  
OY 219 PLPP-----PSMPAVAGAAAGL-----ALLTLGVAGAGAMCWRARRA 256  
DB 904 PRGPRAGNTGAPGSPGVGPRPDAGQPGPKSPGAPGPRGAPPLGLGIGTIGARLAPPG 963  
OY 257 KPSSRRHP-----GSFRGGSLGLG-----GGAMP-----REAPPEGLAL 297  
DB 964 MPGRGSGPQGVGVGSGKPGANGLSGERPPPGQGLPGLAGTAGEDRGNDGSDGLPG 1023  
OY 298 RGG-----GAADPPFCFHYEKVSGDYGPVTVDDGP 330  
DB 1024 RDGSPGKDGKRGNGSPGAP-----GAPGHP-----GFP 1052  
RESULT 43  
CA18\_RABIT

```

ID CA18 RABIT STANDARD; PRT; 744 AA.
AC P14282;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Collagen alpha 1(VIII) chain precursor (Endothelial collagen).
GN COL8A1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89380199; PubMed=2476437;
RA Yamaguchi N., Benya P.D., van der Rest M., Nimniya Y.;
RT "The cloning and sequencing of alpha 1(VIII) collagen cDNAs
RT demonstrate that type VIII collagen is a short chain collagen and
RT contains triple-helical and carboxyl-terminal non-triple-helical
RT domains similar to those of type X collagen."
RL J. Biol. Chem. 264:16022-16029(1989).
CC -1- FUNCTION: MAJOR COMPONENT OF THE DESCENDING MEMBRANE (BASEMENT
CC MEMBRANE) OF CORNEAL ENDOTHELIAL CELLS.
CC -1- SUBUNIT: MAY FORM HOMOTRIMERS, OR HETEROTRIMERS IN ASSOCIATION
CC WITH ALPHA 2(VIII) TYPE COLLAGENS.
CC -1- PM: PROLINES AT THE THIRD POSITION OF THE TRIPLET REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- MISCELLANEOUS: FOUR CONSECUTIVE GLY-PRO-PRO TRIPLETS ARE PRESENT
CC AT THE C-TERMINUS OF THE TRIPLE-HELICAL REGION. THESE MAY PROVIDE
CC HIGH THERMAL STABILITY OF THIS REGION OF THIS REGION.
CC -1- SIMILARITY: STRONG, TO ALPHA 2 TYPES VIII AND X COLLAGENS.
CC -1- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J05042; AAA31204.1; -.
DR FIK; A34246; A34246.
DR InterPro; IPR001073; C1q.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF00386; C1q; 1.
DR Pfam; PF00391; Collagen; 8.
DR PRINTS; PR00007; COMPLEMENTC1Q.
DR SMART; SW00110; C1Q; 1.
DR PROSITE; PS01113; C1Q; 1.
DR Exracellular matrix; Connective tissue; Repeat; Hydroxylation;
KM Glycoprotein; Cell adhesion; Collagen; Signal.
FT SIGNAL 1 20
FT CHAIN 21 744 COLLAGEN ALPHA 1(VIII) CHAIN.
FT DOMAIN 26 117 NONHELICAL REGION (NC2).
FT DOMAIN 118 571 TRIPLE-HELICAL REGION.
FT DOMAIN 572 744 NONHELICAL REGION (NC1).
FT DOMAIN 609 744 C1Q.
SQ SEQUENCE 744 AA; 73358 MW; 2ABCEPFE8274E99 CRC64;

Query Match 7.6%; Score 141; DB 1; Length 744;
Best Local Similarity 23.6%; Pred. No. 0.032;
Matches 83; Conservative 29; Mismatches 108; Indels 132; Gaps 19;

```

```

DB 273 KPGVTFPGPQPLKPPPPGPGPGLGVGVGGPGLPGVGGPQDGIQGGPFGG 332
QY 211 ATSGAGBPLPPSPMPAVA-----GAAGGALLLLGVAGAGAMCRRRR-----A 256
DB 333 KBOQGLPGLPGPGLPGVGGPGLPGPDKD-----RGIGGVPALPBRGKGVGAPGKG 387
QY 257 KPESRHPG-PPSPFGGSLG-----GGGGMGRP-----EAPBELGIALRG--GAAD 304
DB 388 PGEPELPGLPMPMPFGAIGPFGKGGGIVGPGPGPGKGPGLQGFPGKPGFLGVG 447
QY 305 PP--FCPHYKVSDDYGH-----FVYIVQGGP-----QSPNPI 336
DB 448 PGIKGLPPIPKPKKAGKKG.PGLPGVPGLLGPGKGPFGIPQDGLQGPFGI 499

RESULT 44
ID CA18 MOUSE STANDARD; PRT; 743 AA.
AC Q00780; Q9D2V4;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(VIII) chain precursor.
GN COL8A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92362626; PubMed=1499564;
RA Murgaki Y., Shita C., Inoue M., Ooshima A., Olsen B.R.;
RA Nimniya Y.;
RT "Alpha 1(VIII)-collagen gene transcripts encode a short-chain
RT collagen polypeptide and are expressed by various epithelial,
RT endothelial and mesenchymal cells in newborn mouse tissues."
RL Eur. J. Biochem. 207:895-902(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Kidney;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ichii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Saito K., Iizawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleschmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Baren G.,
RA Blake J., Boffelli D., Bojunga N., Carinci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kariya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mommaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
CC -1- FUNCTION: MAJOR COMPONENT OF THE DESCENDING MEMBRANE (BASEMENT
CC MEMBRANE) OF CORNEAL ENDOTHELIAL CELLS.
CC -1- SUBUNIT: MAY FORM HOMOTRIMERS, OR HETEROTRIMERS IN ASSOCIATION
CC WITH ALPHA 2(VIII) TYPE COLLAGENS.
CC -1- TISSUE SPECIFICITY: HIGH LEVELS IN CAVITARY, EYE & SKIN OF
CC NEWBORN MICE. ALSO IN VARIOUS EPITHELIAL, ENDOTHELIAL AND
CC MESENCHYMAL CELLS.
CC -1- PM: PROLINES AT THE THIRD POSITION OF THE TRIPLET REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- SIMILARITY: STRONG, TO ALPHA 2 TYPES VIII AND X COLLAGENS.
CC -1- SIMILARITY: CONTAINS 1 C1Q DOMAIN.

```

```

CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X66976; CA447387.1; -
DR EMBL: X66977; CA447387.1; JOINED.
DR EMBL: AK018742; BAB31383.1; -.
DR PIR: S23779; S23779.
DR WGD: MGI:88463; Col18a1.
DR InterPro: IPR001073; C1q.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF00386; C1q; 1.
DR Pfam: PF01391; Collagen; 7.
DR PRINTS: PRO0007; COMPLEMENTC1Q.
DR SMART: SM00110; C1Q; 1.
DR PROSITE: PS01113; C1Q; 1.
KM Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KM Glycoprotein; Cell adhesion; Collagen; Signal.
FT STGNL 1 28
FT CHAIN 1 28
FT DOMAIN 29 743
FT DOMAIN 118 571
FT DOMAIN 572 743
FT DOMAIN 608 743
FT CONFLICT 6 6
FT CONFLICT 85 85
FT CONFLICT 109 109
FT CONFLICT 248 248
FT CONFLICT 313 313
FT CONFLICT 323 324
FT CONFLICT 361 361
FT CONFLICT 596 596
FT CONFLICT 717 719
SQ SEQUENCE 743 AA; 73454 MW; F584D85BD53897F4 CRC64;

Query Match
Best Local Similarity 21.6%; Score 140.5; DB 1; Length 743;
Matches 101; Conservative 30; Mismatches 135; Indels 209; Gaps 21;

QY 14 GALLLGLVGLVGLSLPEPVNSANKRFQAGGIV---LYQIQIDRLDLCPRARPQ 69
DB 6 GPQLLLGLPILIS-----LNSV-RLIQAGAYYGKPLPPQIPQIPQIPQYQPLG 55
QY 70 ---PH-----SSPNVEFYKLY-----L 83
DB 56 QQVPMPLKDGSLMGKEMPMQYKKEYPPLPYMKIIPPVPMKEVVPKKQKEVPLAS 115
QY 84 VGGAGGRCE---APPAPNLLTCDRPLDLRFTIKFOEYSNLTGHEFRSHDYIITA 139
DB 116 LRGEQGRPRGPRGPRGPR-----PGLFGHMPG-----IX 146
QY 140 TSDGT-----REGLESLOG--GVCLTRGMKYLKVGQSPRGANVRPVSEMPMERDR 190
DB 147 GKRPQGGVPGIGKPGMGKMGKMPGAK-----GEIQPKGEIEMGIPGPQ 196
QY 191 G--AAHSL-----EPKENTLPDPTSNATSGAGPLPPSM----- 225
DB 197 GPRPGHLPRIKPGRGLPGQPGAKG-ERKPRKPGPRGRLQPKKPKGFGMLPGLKLG 255
QY 226 -PAVAGAGLALLLLGVAGAGGAMCWRRRRAXPSRRHPG----- 265
DB 256 PRGHHGPRPGVGLPGVKGPGVTFPRPGQRLKPRGPGQGLIGVGVQPGMPGV 315
QY 266 -----PSFGGSLGAG--GGAGPRB 287
DB 316 GKRGQDGIPIQPGPRGKGQGLPGLPGRGLPGVGKPGPRPKGDRGLGSPVLGPR- 374
QY 288 AEPGEIGIALRGGAADPFPCPHYEKVSGDYGHVYIVDGP-----PQSPF 334

```

```

DB 375 GKXGIGAPMGWG-----PRGEGULPGIRPGMPGPALPGPPKRGKGVVGGPR 425
RESULT 45
ID CA1A HUMAN STANDARD; PRT; 680 AA.
AC 003692;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 1(X) chain precursor.
GN COL10A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92109659; PubMed=1764025;
RA Thomas J.T., Cresswell C.J., Rash B., Nicolai H., Jones T.,
RA Solomon S., Grant M.E., Boot-Handford R.P.;
RT "The human collagen X gene. Complete primary translated sequence and
RT chromosomal localization."
RT Biochem. J. 280:617-623(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93012005; PubMed=1397333;
RA Reichenberger E., Beier F., Luvalle P., Olsen B.R., von der Mark K.,
RA Bertling W.M.;
RT "Genomic organization and full-length cDNA sequence of human collagen
RT X."
RT FEBS Lett. 311:305-310(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX Beier F., Lamm M.B., von der Mark K.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX Williams S.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 52-680 FROM N.A.
RX MEDLINE=92267014; PubMed=1587271;
RA Ape S.S., Seldin M.F., Hayashi M., Olsen B.R.;
RT "Cloning of the human and mouse type X collagen genes and mapping of
RT the mouse type X collagen gene to chromosome 10."
RT Eur. J. Biochem. 206:217-224(1992).
RN [6]
RP SEQUENCE OF 561-666 FROM N.A.
RX MEDLINE=91243836; PubMed=2037056;
RA Ape S., Mattei M.-G., Olsen B.R.;
RT "Cloning of human alpha 1(X) collagen DNA and localization of the
RT COL10A1 gene to the q21-q22 region of human chromosome 6."
RT FEBS Lett. 282:393-396(1991).
RN [7]
RP SEQUENCE OF 547-655 FROM N.A.
RX MEDLINE=92077285; PubMed=1743401;
RA Reichenberger E., Aigner T., von der Mark K., Stoeck H., Bertling W.;
RT "In situ hybridization studies on the expression of type X collagen
RT in fetal human cartilage."
RT Dev. Biol. 148:562-572(1991).
RN [8]
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kilvianem H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels."
RT Hum. Mutat. 9:300-315(1997).
RN [9]
RP VARIANTS SWCD ASP-598 AND PRO-614.
RX MEDLINE=9416476; PubMed=8304336;

```

RA Wallis G.E., Raab B., Sweetman W.A., Thomas J.T., Super M., Evans G.,  
 RA Grant M.E., Booc-Handford R.P.,  
 RT "Amino acid substitutions of conserved residues in the  
 RT carboxyl-terminal domain of the alpha 1(X) chain of type X collagen  
 RT occur in two unrelated families with metaphyseal chondrodysplasia  
 RT type Schmid." ;  
 RL Am. J. Hum. Genet. 54:169-178(1994).  
 RN [10]  
 RP VARIANT SMCD ARG-591.  
 RX MEDLINE=94272470; PubMed=8004095.  
 RA McIntosh I., Abbott M.H., Waxman M.L., Olsen B.R., Francomano C.A.,  
 RT "Additional mutations of type X collagen confirm COL10A1 as the  
 RT Schmid metaphyseal chondrodysplasia locus." ;  
 RL Hum. Mol. Genet. 3:303-307(1994).  
 RN [11]  
 RP VARIANT SMCD VAL-618.  
 RX MEDLINE=95181449; PubMed=7876225.  
 RA Chan D., Cole W.G., Rogers J.G., Bateman J.F.,  
 RT "Type X collagen multimer assembly in vitro is prevented by a Gly618  
 RT to Val mutation in the alpha 1(X) NC1 domain resulting in Schmid  
 RL metaphyseal chondrodysplasia." ;  
 RL J. Biol. Chem. 270:4556-4562(1995).  
 RN [12]  
 RP VARIANTS SMCD ARG-545; GLU-555; HIS-597; LYS-617; ARG-644 AND GLY-648  
 RX MEDLINE=95331767; PubMed=7607655.  
 RA Bonaventure J., Chaminate F., Maroteaux P.,  
 RT "Mutations in three subdomains of the carboxy-terminal region of  
 RT collagen type X account for most of the Schmid metaphyseal  
 RL dysplasias." ;  
 RL Hum. Genet. 96:58-64(1995).  
 RN [13]  
 RP VARIANT SMCD PRO-600.  
 RX MEDLINE=96375754; PubMed=8782043;  
 RA Wallis G.A., Raab B., Sykes B., Bonaventure J., Maroteaux P.,  
 RA Zabel B., Wynne-Davies R., Grant M.E., Booc-Handford R.P.,  
 RT "Mutations within the gene encoding the alpha 1(X) chain of type X  
 RT collagen (COL10A1) cause metaphyseal chondrodysplasia type Schmid but  
 RT not several other forms of metaphyseal chondrodysplasia." ;  
 RL J. Med. Genet. 33:450-457(1996).  
 RN [14]  
 RP VARIANTS SMCD GLU-18 AND ARG-18.  
 RX MEDLINE=97220591; PubMed=9067755.  
 RA Ikegawa S., Nakamura K., Nagano A., Haga N., Nakamura Y.,  
 RT "Mutations in the N-terminal globular domain of the type X collagen  
 RT gene (COL10A1) in patients with schmid metaphyseal  
 RL chondrodysplasia." ;  
 RL Hum. Mutat. 9:131-135(1997).  
 RN [15]  
 RP VARIANTS SMD GLU-595.  
 RX MEDLINE=99057503; PubMed=8637818;  
 RA Ikegawa S., Nishimura G., Nagai T., Hasegawa T., Onashi H.,  
 RA Nakamura Y.,  
 RT "Mutation of the type X collagen gene 'COL10A1' causes  
 RT spondylometaphyseal dysplasia." ;  
 RL Am. J. Hum. Genet. 63:1659-1662(1998).  
 RN [16]  
 RP VARIANT SMCD CYS-597.  
 RX MEDLINE=99069781; PubMed=9852679;  
 RA Sawai H., Ida A., Nakata Y., Koyama K.,  
 RT "Novel missense mutation resulting in the substitution of tyrosine by  
 RT cysteine at codon 597 of the type X collagen gene associated with  
 RT Schmid metaphyseal chondrodysplasia." ;  
 RL J. Hum. Genet. 43:259-261(1998).  
 CC -1- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERPHOSPHIC  
 CC CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE  
 CC MINERALIZATION ZONES OF HYALINE CARTILAGE.  
 CC -1- SUBUNIT: HOMOOLIMER.  
 CC -1- PM: PROLINS ARE AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- DISEASE: DEFECTS IN COL10A1 ARE THE CAUSE OF SCHMID TYPE  
 CC METAPHYSEAL CHONDRODYSPLASIA (SMCD), A DOMINANTLY INHERITED  
 CC DISORDER OF THE OSSEOUS SKELETON. THE CARDINAL FEATURES OF THE  
 CC PHENOTYPE ARE MILD SHORT STATURE, COXA VARA AND A WADDLING GAIT.

```
CC METAPHOSARY USUALLY SHOWS SCLEROSIS OF THE RIMS, FLARING OF THE  
CC METAPHYSESS, AND A WIDE IRREGULAR GROWTH PLATE, ESPECIALLY OF THE  
CC KNEES.  
CC -1- DISEASE : DEFECTS IN COL0A1 ARE THE CAUSE OF SPONDYLOMETAPHYEAL  
CC DYSPLASIA (SMD). SMD COMPRISES A HETEROGENEOUS GROUP OF HERITABLE  
CC SKELETAL DYSPLASIAS CHARACTERIZED BY MODIFICATIONS OF THE  
CC VERTEBRAL BODIES OF THE SPINE AND METAPHYSSES OF THE TUBULAR BONES.  
CC -1- SIMILARITY : STRONG TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.  
CC -1- SIMILARITY : CONTAINS 1 C1Q DOMAIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboratio  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation o  
CC the European Bioinformatics Institute. There are no restrictions on its u  
CC use by non-profit institutions as long as their content is in no way co  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.ebi.ac.uk/announce/  
CC or send an email to license@ebi.ac.uk).
```

```
CC -----  
DR EMBL/X60382; CAA42933.1; -;  
DR EMBL/X65120; CAA46236.1; -;  
DR EMBL/X98568; CAA67178.1; -;  
DR EMBL/A1121963; CAB87590.1; -;  
DR EMBL/S68531; AAC60615.1; -;  
DR EMBL/X58879; CAA41686.1; -;  
DR EMBL/M74050; AAA61221.1; -;  
DR EMBL/X72579; CAA51170.1; -;  
DR EMBL/X72580; CAA51170.1; JOINED.  
DR PIR/S15826; S15826.  
DR PIR/S30086; S30086.  
DR PIR/A43901; A43901.  
DR PIR/S18249; S18249.  
DR PIR/S21856; S21856.  
DR PIR/S26396; S26396.  
DR Genew/HGNC:2185; COL10A1.  
DR MIM/120110; -.  
DR MIM/156500; -.  
DR MIM/184250; -.  
DR InterPro/IPR001073; C1q.  
DR InterPro/IPRO00087; Collagen.  
DR Pfam/PF00386; C1q; 1.  
DR Pfam/PF01391; Collagen; 8.  
DR PRINTS/PR00007; COMPLEMENTC1Q.  
DR SMART/SMO0110; C1Q_1.  
DR PROSITE/PS01113; C1q; 1.  
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Cartilage; Collagen; Signal; Disease mutation; Polymorphism.  
FT SIGNAL      1          18  
FT CHAIN       19         680  
FT DOMAIN     19         56  
FT            57        519  
FT           520        680  
FT DOMAIN     545        680  
FT VARIANTS    18         18  
FT VARIANT     18         18  
FT             545        545  
FT VARIANTS    591        591  
  
Query Match               7.6%; Score 140; DB 1; Length 680;  
Best Local Similarity   29.1%; Pred. No. 0.034;  
Matches 57; Conservative 13; Mismatches 84; Indels 42; Gaps 9
```

```
Oy 171 PRG---CAVRRKP-----VSMPMERRDGAHSLFKENTLDG-----DPTNSA 211  
Db 70 PRGHDPGGPGPKGXGYSGPLQGEGFLGPFGPSPAVGVGGLPGKGRGPPYGPKDV 129  
Cy 212 TSRGAGGLELPSPMAVAAGAAGJALLLLGYAAGAACMCRRAKRPAESSRHG-PGSFG 270  
Cb 130 GPAGLPGRGPGPFPGIGFPAG--ISVPAGPGQQGSPGACGPRGFPEBKAKPVGMANG 186
```

QY 271 RGGSLGIGGGGNGPREAPGELGALRGG--GAADP-----FCPHYKSGDYGH 320  
 DB 187 QNGEMGYGAPG-----RPGRGJLPGPGPGPGSPGVGRGNGVPGCGIRGNDGAF 239  
 QY 321 PVIYVDPG--PQSP 334  
 DB 240 PGMGPBGPGPGPP 255

RESULT 46  
 CAC2\_HABCO STANDARD; PRT; 210 AA.  
 ID CAC2\_HABCO  
 AC P16252;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cuticle collagen 2C (Fragment).  
 GN 2C.  
 OS Haemonchus contortus (Barber pole worm).  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;  
 OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.  
 RX NCBI\_TaxID=6289;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90136718; PubMed=2615789;  
 RA Shanahan L.M., Pratt D., Bolavenue R.J., Cox G.N.;  
 RT "Cuticle collagen genes of Haemonchus contortus and Caenorhabditis  
 elegans are highly conserved."  
 RL Mol. Biochem. Parasitol. 37:73-86(1989).  
 CC -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
 CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
 CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.  
 CC -1- MISCELLANEOUS: THIS PROTEIN SHOWS 4 POTENTIAL TRIPLE-HELICAL  
 CC REGIONS, WHICH CONTAIN GLYCINE AS EVERY THIRD AMINO ACID.  
 CC -1- MISCELLANEOUS: IN ALL NEMATODE CUTICLE COLLAGENS, THE POLYPEPTIDE  
 CC CHAINS ARE COMPLEXED WITHIN THE CUTICLE BY DISULFIDE BONDS AND  
 CC OTHER TYPES OF COVALENT CROSS-LINKS.  
 CC -1- SIMILARITY: BELONGS TO THE CUTICULAR COLLAGEN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; J04670; AAA29172.1;  
 CC InterPro; IPR000087; Collagen.  
 CC Pfam; PF01391; Collagen 3.  
 CC Cuticle; Connective tissue; Repeat; Multigene family; Collagen.  
 CC NON TER 1  
 CC FT 1  
 CC SQ SEQUENCE 210 AA; 19562 MW; E15FA9A2DF3D74B CRC64;

Query Match 7.5%; Score 139.5; DB 1; Length 210;  
 Best Local Similarity 31.1%; Pred. No. 0.011; Indels 31; Gaps 5;  
 Matches 50; Conservative 5; Mismatches 75;

QY 176 VRRKVSSEMPERDRGAHSLRPGKENTLPGDP--TSNATSRGAEPLPPSPMPAVAGAAG 233  
 DB 51 RIIPKCPQGRPPGPPGPGPGTGNAGANAPAPPPGPKPPGPPGAGAP- 109  
 QY 234 GLALLILGVAGAGMCMRRRAKPSSESRHPG---PSFGRGSLGIGGGGNGPREAE 289  
 DB 110 -----GAAGPAGANA-----PSEPLVPGPPPGPTGEGPGPNAGAPGAGAGA 155  
 QY 290 PSELGIALRGGAADPPFCPHYKSGDYGPVYIVQDGP 330  
 DB 156 PSEKGRGGDGHGAP-----GNAHGCGGCGGPP 186

ID CAlB MOUSE STANDARD; PRT; 1804 AA.  
 AC Q6125; O64047;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 1(XI) chain precursor.  
 GN COL11A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP TISSUE=Embryo;  
 RX MEDLINE=96015067; PubMed=8530046;  
 RA Yoshioke H., Inoguchi K., Knaletuzzaman M., Ninomiya Y.,  
 RA Andrikopoulos K., Ramirez F.;  
 RT "Coding sequence and alternative splicing of the mouse alpha 1(XI)  
 RT collagen gene (Col11a1)."  
 RL Genomics 28:337-340(1995).  
 RN [2]  
 RP SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.  
 RC STRAIN=C57BL/6;  
 RX MEDLINE=95163095; PubMed=7859283;  
 RA Li Y., Lacerda D.A., Warman M.L., Belier D.R., Yoshioke H.,  
 RA Ninomiya Y., Oxford J.T., Morris N.P., Andrikopoulos K.,  
 RA Ramirez F., Wardell B.B., Liffert G.D., Tetscher C., Woodward S.R.,  
 RA Taylor B.A., Seemuller R.E., Olsen B.R.;  
 RT "A fibrillar collagen gene, Col1a1, is essential for skeletal  
 RT morphogenesis."  
 RL Cell 80:423-430(1995).  
 CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY  
 CC CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.  
 CC -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),  
 CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL  
 CC MODIFICATION OF ALPHA 1(XI). ALPHA 1(XI) CAN ALSO BE FOUND INSTEAD  
 CC OF ALPHA 3(XI)=1(XI) (BY SIMILARITY).  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- DISEASE: DEFECTS IN COL1A1 ARE ASSOCIATED WITH CHONDRODYSPLASIA,  
 CC AN AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY SKELETAL DEFECTS  
 CC CAUSED BY ABNORMALITIES IN THE CARTILAGE OF LIMBS, RIBS, MANDIBLES  
 CC AND TRACHEA.  
 CC -1- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.  
 CC -1- SIMILARITY: HIGH TO ALPHA 1(V) AND ALPHA 3(V) CHAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D38162; BAA07367.1;  
 CC EMBL; S74574; AAB33439.1;  
 CC MGI; MGI:88446; Col11a1.  
 CC InterPro; IPR000087; Collagen.  
 CC InterPro; IPR000085; Fib collagen\_C.  
 CC InterPro; IPR001791; LamInG.  
 CC InterPro; IPR001230; Premyl\_site.  
 CC InterPro; IPR003129; TSPN.  
 CC Pfam; PF01391; Collagen; 16.  
 CC Pfam; PF01410; COLF1; 1.  
 CC Pfam; PF02210; TSPN; 1.  
 CC ProDom; PD000007; Collagen; 1.  
 CC ProDom; PD002078; Fib collagen\_C; 1.  
 CC SMART; SM00038; COLF1; 1.  
 CC SMART; SM00282; LamG; 1.  
 CC SMART; SM00210; TSPN; 1.  
 CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;



KM Glycoprotein; Collagen; Signal; Alternative splicing;  
 FT Disease mutation. 35  
 FT SIGNAL 1  
 FT PROPEP 511 POTENTIAL.  
 FT CHAIN 312 AMINO-TERMINAL PROPEPTIDE (POTENTIAL).  
 FT PROPEP 1561 COLLAGEN ALPHA 1(XI) CHAIN.  
 FT PROPEP 1562 CARBOXYL-TERMINAL PROPEPTIDE.  
 FT DOMAIN 37 NONHELIICAL REGION.  
 FT DOMAIN 418 TRIPLE-HELICAL REGION (INTERRUPTED).  
 FT DOMAIN 506 SHORT NONHELIICAL SEGMENT.  
 FT DOMAIN 510 TRIPLE-HELICAL REGION.  
 FT DOMAIN 527 TRIPLE-HELICAL REGION.  
 FT DOMAIN 540 NONHELIICAL REGION (C-TERMINAL).  
 FT DOMAIN 1541 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 1638  
 FT SITE 610  
 FT SITE 1450 CROSSLINKING.  
 FT VARSPLIC 329 MISSING (IN SHORT ISOFORM).  
 FT VARIANT 189 LRSERS -> SIEVDR (IN CHONDRODYSPLASIA).  
 FT VARIANT 196 MISSING (IN CHONDRODYSPLASIA).  
 SQ SEQUENCE 1804 AA; 180963 MW; FEZDB9DED1E4219A CRC64;

Query Match 7.5%; Score 139; DB 1; Length 1804;  
 Best Local Similarity 25.7%; Pred. No. 0.1; Mismatches 170; Gaps 24;  
 Matches 105; Conservative 17; Indels 170; Gaps 24;

QY 2 GPP-HSGPGGVVAGALLLGLVGL--VSGLSLEPVYNSANKRQAGGYLYPQIGDRL 58  
 DB 440 GPPGPNAGAG-----LMGPPGLQSPSLRGD-----GDR- 469  
 QY 59 DLCCPRAPPPGPHSSPNVEFYKLYLVGAGGRCRCPAPNLLITCDPDLIRF----- 113  
 DB 470 -----GPPRPPGLP-----GADG--LPPPGTMTM-----LPRYGGDGS 502  
 QY 114 ---TIKPGVEYPMNGHFHSHNDYIIATSDGREGLESIGGVCLTRGKTKLRYG-- 166  
 DB 503 KGPITISQEOAQAQ-----ALIQAKIALRGPPEPWGLTG-----RGPV 541  
 QY 169 -----QSPRGAVPRKPYSE-----WPMER-----DRG-- 191  
 DB 542 GGPSTGAKGSGDPGPGGPRGPGPTGPKGRPGADGGRGKPGSGSKDGRGFD 601  
 QY 192 -----AAHSLRPGKENLPDPTSNATSRGAGCPPPSPMPVAAAGALALLIGVA 243  
 DB 602 GLPLPGDKGRHGRGPGPGLPGDDG--MKGDEIGPGLPEAAPRG-----LLGPR 655  
 QY 244 GA-----GGAMCWRRRRAXPSESHPG--PQSPRGSGSLG--GGGGGMPREAPGE 292  
 DB 656 GTPGPPGPGIGIGIDGPGPKMNGPQGPBPQGGNPGSGSLGPGQGIIGP-----PGE 711  
 QY 293 LGALRGS-----GAADPPCFPHYKXVSGDYGHFVTVQGP--PQSP 333  
 DB 712 KGPGKGPLAGLPAGADGPPGHGPKRGSGGEXG-----ALGPPGPGP 753

RESULT 48  
 CA12\_MOUSE STANDARD; PRT; 1459 AA.  
 AC P28481;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 1(II) chain precursor [contains: Chondrocalcin].  
 GN COL2A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 R2 SEQUENCE FROM N.A.  
 RK MEDLINE=91358489; PubMed=1885613;  
 RA Metzaranta M., Toman D., de Crombrughe B., Vuorio E.;  
 RT "Mouse type II collagen gene. Complete nucleotide sequence, exon  
 structure, and alternative splicing.";  
 RL J. Biol. Chem. 266:16862-16869 (1991).

RN [2]  
 RP SEQUENCE OF 1455-1459 FROM N.A.  
 RX MEDLINE=91274355; PubMed=2054384;  
 RA Metzaranta M., Toman D., de Crombrughe B., Vuorio E.;  
 RT "Specific hybridization probes for mouse type I, II, III and IX  
 collagen mRNAs.";  
 RL Biochim. Biophys. Acta 1089:241-243 (1991).  
 CC -1- FUNCTION: COLLAGEN TYPE II IS SPECIFIC FOR CARTILAGINOUS TISSUES.  
 CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(II) CHAINS.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPLEPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: CONTAINS 1 VWF C DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M65161; AAA68100.1; -;  
 CC EMBL; X57982; CAA41047.1; -;  
 CC MGD; MGI:88452; Col2a1.  
 CC InterPro; IPR000087; Collagen.  
 CC InterPro; IPR000885; Fib collagen\_C.  
 CC InterPro; IPR001007; VWF\_C.  
 CC Pfam; PF00093; VWC; 1.  
 CC Pfam; PF01391; Collagen; 18.  
 CC Pfam; PF01410; COLFI; 1.  
 CC ProDom; PD000007; Collagen; 3.  
 CC ProDom; PD002078; Fib collagen\_C; 1.  
 CC SMART; SMO0214; COLFI; 1.  
 CC SMART; SMO0214; VWC; 1.  
 CC PROSITE; PS01208; VWF; 1.  
 CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KM Glycoprotein; Collagen; Cartilage; Signal; Alternative splicing.  
 FT SIGNAL 1  
 FT PROPEP 26  
 FT CHAIN 154  
 FT PROPEP 1214  
 FT PROPEP 1214  
 FT DOMAIN 32  
 FT DOMAIN 173  
 FT DOMAIN 1187  
 FT VARSPLIC 29  
 FT VARSPLIC 30  
 FT VARSPLIC 98  
 SQ SEQUENCE 1459 AA; 139154 MW; F6C84FA7C532E7F2 CRC64;

Query Match 7.5%; Score 138.5; DB 1; Length 1459;  
 Best Local Similarity 30.1%; Pred. No. 0.09;  
 Matches 58; Conservative 13; Mismatches 75; Indels 47; Gaps 9;

QY 173 GGAUVRKPYSEMPMRDGAHSLRPGKENT-----PGD--PISNATSRGAEGLPPSP 224  
 DB 470 GAGPRTGPPGGERGAGNGRF-----PGDGLAGPAGGEGSGSLAPKANGPPGPG 524  
 QY 225 MPAYVGAAGGLALLLGVAGAGAMCWRRRRAXPSESHPG-----GSPRGSGSLGSGG 280  
 DB 525 EPGLGARG-----LTGRPGDAGPGKXVPSGAPBEDRPPGPGGARGGPGVWVGP 579  
 QY 281 GGMGPREAPGELGTA-----LRG-----GAADPPCFPHYKXVSGDYGHFVTVQGP 329  
 DB 580 KGANBPPGAKGKGLAGAPGLRGLPGKDGETGAAGPPGSPGAPGRGQAGP-----GP 633  
 QY 330 -----PQSP 334  
 DB 634 SGFGLPBPBPBP 646

RESULT 49  
 CA2B\_HUMAN

ID CA2B\_HUMAN STANDARD; PRT; 1736 AA.  
 AC P13942; Q13273; Q13271; Q13272; Q07751; Q99866; Q9UIP9;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 2(XI) chain precursor.  
 GN COL11A2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96032717; PubMed=7559422;  
 RA Vuorio M.M., Pihlajamaa T., Vandenbergh P., Prockop D.J.,  
 RA Ala-Kokko L.;  
 RT "The human COL11A2 gene structure indicates that the gene has not  
 RT evolved with the genes for the major fibrillar collagens.";  
 RL J. Biol. Chem. 270:22873-22881(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Tubby B.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 59-807 FROM N.A.  
 RC TISSUE=Cartilage;  
 RX MEDLINE=93314795; PubMed=8325374;  
 RA Zhidkova N.I., Brewton R.G., Wayne R.;  
 RA "Molecular cloning of PARP (proline/arginine-rich protein) from human  
 RT cartilage and subsequent demonstration that PARP is a fragment of the  
 RT NH2-terminal domain of the collagen alpha 2(XI) chain.";  
 RL Nucleic Acids Res. 21:525-528(1993).  
 RN [4]  
 RP SEQUENCE OF 730-1690 FROM N.A.  
 RX MEDLINE=89340485; PubMed=2760050;  
 RA Kimura T., Cheah K.S.E., Chan S.D.H., Lui V.C.H., Mattei M.-G.,  
 RA van der Rest M., Ono K., Solomon E., Ninomiya Y., Olsen B.R.;  
 RT "The human alpha 2(XI) collagen (COL11A2) chain. Molecular cloning of  
 RT cDNA and genomic DNA reveals characteristics of a fibrillar collagen  
 RT with differences in genomic organization.";  
 RL J. Biol. Chem. 264:13910-13916(1989).  
 RN [5]  
 RP SEQUENCE OF 1-537 FROM N.A.  
 RX MEDLINE=96435918; PubMed=8838804;  
 RA Lui V.C., Ng L.J., Sat E.W., Cheah K.S.;  
 RT "The human alpha 2(XI) collagen gene (COL11A2): completion of coding  
 RT information, identification of the promoter sequence, and precise  
 RT localization within the major histocompatibility complex reveal  
 RT overlap with the KES gene.";  
 RL Genomics 32:401-412(1996).  
 RN [6]  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE=95238468; PubMed=7721876;  
 RA Zhidkova N.I., Justice S.K., Wayne R.;  
 RT "Alternative mRNA processing occurs in the variable region of the  
 RT pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains.";  
 RL J. Biol. Chem. 270:9486-9493(1995).  
 RN [7]  
 RP DISEASE.  
 RX PubMed=10677296;  
 RA Melkonian M., Brunner H.G., Manouvrier S., Hennekam R.,  
 RA Superti-Furga A., Kaeaeiaenen H., Pauli R.M., van Essen T.,  
 RA Warman M.L., Bonaventura J., Mlyn P., Ala-Kokko L.;  
 RT "Autosomal recessive osteopoddyomegaepiphyseal dysplasia is  
 RT associated with loss-of-function mutations in the COL11A2 gene.";  
 RL Am. J. Hum. Genet. 66:368-377(2000).  
 RN [8]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=97253959; PubMed=9101290;  
 RA Kuitaniemi H., Tromp G., Prockop D.J.;  
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
 RT associated collagen (type IX), and network-forming collagen (type X)  
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";

RL Hum. Mutat. 9:300-315(1997).  
 RN [9]  
 RP VARIANT OSMED ARG-661.  
 RX MEDLINE=95163096; PubMed=7859284;  
 RA Vilkki M., Mariman E.C.M., Lui V.C.H., Zhidkova N.I., Tiller G.E.,  
 RA Golding M.B., van Beersum S.B.C., de Waal Malefijt W.C.,  
 RA van den Hoogen F.H.J., Roberts H.-H., Wayne R., Cheah K.S.E.,  
 RA Olsen B.R., Warman M.L., Brunner H.G.;  
 RT "Autosomal dominant and recessive osteochondrodysplasias associated  
 RT with the COL11A2 locus.";  
 RL Cell 80:431-437(1995).  
 RN [10]  
 RP VARIANTS GLY-593; LYS-824; LEU-879; THR-1316 AND GLN-1600.  
 RX PubMed=9585596;  
 RA Koga H., Sakou T., Taketomi E., Hayashi K., Numasawa T., Harata S.,  
 RA Yone K., Matsunaga S., Ohterud B., Inoue I., Leppert M.;  
 RT "Genetic mapping of ossification of the posterior longitudinal  
 RT ligament of the spine.";  
 RL Am. J. Hum. Genet. 62:1460-1467(1998).  
 RN [11]  
 RP VARIANT W25 GLU-955.  
 RX PubMed=9805126;  
 RA Pihlajamaa T., Prockop D.J., Faber J., Winterpacht A., Zabel B.,  
 RA Giedion A., Wiesbauer P., Spranger J., Ala-Kokko L.;  
 RT "Heterozygous glycine substitution in the COL11A2 gene in the original  
 RT patient with the Weissenbacher-Zweymüller syndrome demonstrates its  
 RT identity with heterozygous OSMD (homocystin-like syndrome).";  
 RL Am. J. Med. Genet. 80:115-120(1998).  
 RN [12]  
 RP VARIANT STU3 940-GLY--PRO-948 DEL.  
 RX PubMed=9506662;  
 RA Strick-Osadek D.A., Murray M.A., Scott J.A., Lavery M.A., Warman M.L.,  
 RA Rodin N.H.;  
 RT "Stickler syndrome without eye involvement is caused by mutations in  
 RT COL11A2, the gene encoding the alpha-2(XI) chain of type XI  
 RT collagen.";  
 RL J. Pediatr. 132:368-371(1998).  
 RN [13]  
 RP VARIANTS DPM13 GLU-808 AND CYS-1034, AND REVISIONS TO 1031-1032.  
 RX MEDLINE=20047768; PubMed=10581026;  
 RA McGulirt W.T., Prasad S.D., Griffith A.J., Kunst H.P.M., Green G.E.,  
 RA Shargel K.B., Runge C., Huybrechts C., Mueller R.F., Lynch E.,  
 RA King M.-C., Brunner H.G., Creemers C.W.R.J., Takano M., Li S.-W.,  
 RA Arita M., Wayne R., Prockop D.J., Van Camp G., Smith R.J.H.;  
 RT "Mutations in COL11A2 cause non-syndromic hearing loss (DPM13).";  
 RL Nat. Genet. 23:413-419(1999).  
 RN [14]  
 RP FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY  
 CC CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.  
 CC SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),  
 CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL  
 CC MODIFICATION OF ALPHA 1(XI). ALPHA 3(XI) CAN ALSO BE FOUND INSTEAD  
 CC OF ALPHA 3(XI)=1(II).  
 CC [15]  
 CC -1- ALTERNATIVE PRODUCTS: 8 isoforms; 1 (shown here), 2, 3, 4, 5, 6, 7  
 CC and 8; may be produced by alternative splicing. They lack exons 6,  
 CC 7 or 8 or a combination of these exons.  
 CC -1- PPM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- PPM: A DISULFIDE-BONDED PEPTIDE CALLED PROLINE/ARGININE-RICH  
 CC PROTEIN OR PARP IS RELEASED FROM THE AMINO TERMINUS DURING  
 CC EXTRACELLULAR PROCESSING AND IS SUBSEQUENTLY RETAINED IN THE  
 CC EXTRACELLULAR MATRIX FROM WHICH IT CAN BE ISOLATED IN SIGNIFICANT  
 CC AMOUNTS.  
 CC -1- DISEASE: Defects in COL11A2 are the cause of Stickler syndrome  
 CC type 3 (ST3). It is an autosomal dominant disorder characterized  
 CC by oro-facial, auditory and skeletal manifestations, such as  
 CC midfacial hypoplasia, cleft palate, osteoarthritis, and  
 CC sensorineural hearing loss. Differently from Stickler syndrome  
 CC type 1 and 2, no ocular involvement is observed. This disorder is  
 CC also referred to as Stickler-like syndrome or non-ocular Stickler  
 CC syndrome.  
 CC -1- DISEASE: Defects in COL11A2 are the cause of autosomal recessive  
 CC osteopoddyomegaepiphyseal dysplasia (OSMED), a skeletal dysplasia  
 CC accompanied by severe hearing loss. The phenotype overlaps that of

```

CC autosomal dominant skeletal disorders (Stickler and Marshall
CC syndromes) but can be distinguished by disproportionately short
CC limbs and lack of ocular involvement.
CC -1- DISEASE: Defects in COL1A2 are the cause of Weissenbacher-
CC Zeymüller syndrome (WZS), an autosomal dominant disorder allelic
CC with STL3 and OSMD. It is also referred to as heterozygous OSMD.
CC -1- DISEASE: Defects in COL1A2 are the cause of autosomal dominant
CC nonsyndromic sensorineural deafness type 13 (DFNA13). Affected
CC individuals experience progressive hearing loss beginning in the
CC second to fourth decades, eventually making use of amplification
CC mandatory.
CC -1- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.
CC -1- DATABASE: NAME=Hereditary hearing loss homepage;
CC NOTE=Gene page;
CC WWW="http://www.uia.ac.be/dnalab/hnh/hnhgenes.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32169; AAC50214.1; -
CC EMBL; U32169; AAC50213.1; -
CC EMBL; U32169; AAC50215.1; -
CC EMBL; AL031228; CAA30240.1; -
CC EMBL; L19887; AAA35498.1; -
CC EMBL; U04874; AAA52034.1; -
CC EMBL; U41069; AAC17464.1; -
CC EMBL; U41065; AAC17464.1; JOINED.
CC EMBL; U41065; AAC17464.1; JOINED.
CC EMBL; U41067; AAC17464.1; JOINED.
CC PIR; A32645; A32645.
CC Genew; HGNC:2187; COL1A2.
CC MIM; 120290; -
CC MIM; 184840; -
CC MIM; 215150; -
CC MIM; 277610; -
CC MIM; 601868; -
CC RX INTERPRO: IPR000087; Collagen.
CC RX INTERPRO: IPR000885; Fib_collagen_C.
CC RX INTERPRO: IPR001791; Laminin_G.
CC RX INTERPRO: IPR001230; Prenyl_site.
CC RX INTERPRO: IPR003129; TSPN.
CC PFam; PF01391; Collagen; 18.
CC PFam; PF01410; COLFR; 1.
CC PFam; PF02210; TSPN; 1.
CC Pfam; PF000007; Collagen; 1.
CC Pfam; PF000078; Fib_collagen_C; 1.
CC SMART; SM00038; COLFR; 1.
CC SMART; SM00282; LamG; 1.
CC SMART; SM00210; TSPN; 1.
CC
CC Query Match 7.5%; Score 138.5; DB 1; Length 1736;
CC Best Local Similarity 31.7%; Pred. No. 0.11; Indels 49; Gaps 14;
CC Matches 64; Conservative 18; Mismatches 71;

```

```

DB 1482 KBEKVGQPPGHP-----GPP 1497
RESULT 50
CA24 CAEEL STANDARD; PRT; 1758 AA.
ID CA24 CAEEL
PI 17140; Q19098; Q19099;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 2(IV) chain precursor (lethal protein 2).
GN LET-2 OR CLB-1 OR F01G12.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC STRAIN=Bristol N2;
RX MEDLINE=94012964; PubMed=7691828;
RA Sibley M.H., Johnson J.J., Mello C.C., Kramer J.M.;
RT "Genetic identification, sequence, and alternative splicing of the
RT Caenorhabditis elegans alpha 2(IV) collagen gene.";
RL J. Cell Biol. 123:255-264(1993).
RN [2]
RP PRELIMINARY SEQUENCE OF 1495-1758 FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=90008929; PubMed=2793871;
RA Guo X., Kramer J.M.;
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen
RT genes are located on separate chromosomes.";
RL J. Biol. Chem. 264:17574-17582(1989).
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=Bristol N2;
RA Wu X., Le T.T.;
RL Submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.
RN [4]
RP VARIANTS.
RX MEDLINE=94320591; PubMed=8045258;
RA Sibley M.H., Graham P.L., von Mendel N., Kramer J.M.;
RT "Mutations in the alpha 2(IV) basement membrane collagen gene of
RT Caenorhabditis elegans produce phenotypes of differing severities.";
RL EMO J. 13:3278-3285(1994).
CC -1- FUNCTION: Collagen type IV is specific for basement membranes.
CC Vital for embryonic development.
CC -1- SUBUNIT: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.
CC TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH
CC INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NCI
CC DOMAINS.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms, I/a (shown here) and II/b; are
CC produced by alternative splicing.
CC -1- DEVELOPMENTAL STAGE: Isoform I is predominant in embryos and
CC isoform II is predominant in the larvae and adults.
CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC TRIPLE-HELICAL 7S DOMAIN.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC IV COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

